

14:26:18 2004

ug-09-501-171a-4.rai

IBM Compatible
SYSTEM: Windows 95
FastSEQ for Windows Version 2.0b
ICATION DATA:
NUMBER: US/08/982.785A

3.2%; Score 206; DB 3; Length 39;
 larity 100.0%; Pred. No. 2.3e-08;
 conservative 0; Mismatches 0; Indels

KPYLTLLSECSNPDTLEGAAGALONLAAGSWKWSVY 883
 |||||
 KPYLTLLSECSNPDTLEGAAGALONLAAGSWKWSVY 39
 |||||

lication US/08982785A

929
ATTENTION:
Kosik, Kenneth S.
Zhou, Jianhua
ENTENTION: ALARM RELATED PEPTIDES AND
ENTENTION: NUCLEIC ACIDS AND DIAGNOSIS USING THEM
QUENCES: 13
CE ADDRESS:
Fish & Richardson P.C.
25 Franklin Street
ton

```

USA
0-2804
DABLE FORM:
E: Diskette
IBM Compatible
SYSTEM: Windows 95
FASTSQL for Windows Version 2.0b
ICATION DATA:
N NUMBER: US/08/982.785A
E: 02-DEC-1997

```

E: 02-DEC-1996
NT INFORMATION:
eman, John W.
ON NUMBER: 29,000
DOCKET NUMBER:
ATTN INFORMATION:
617/542-5070
617/542-8906
0154
R SEQ ID NO: 5:
RACTERISTICS:
l amino acids

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-982-785A-5

```

Query Match 3.2%; Score 206; DB 3; Length 41;
Best Local Similarity 100.0%; Pred. No. 2.4e-08;
Matches 41: Conservative 0; Mismatches 0; Indels 0;

QY 636 GGIPALVRLLRKTTDLEIRELVTGVLWNLSDDLKMPPIQ 676

dp 1 GGIPALVRLLRKTTDLEIRELVTGVLWNLSDDLKMPPIQ 41

```

RESULT 14
US-09-345-473E-40
; Sequence 40, Application US/09345473E
; Patent No. 6558903
; GENERAL INFORMATION:
; APPLICANT: Hodge, Martin
; TITLE OF INVENTION: No. 6558903el Kinases
; FILE REFERENCE: 35800/183781
; CURRENT APPLICATION NUMBER: US/09/345,473E
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 1601
; TYPE: PRT
; ORGANISM: C. elegans
US-09-345-473E-40

```

Query Match 3.1%; Score 197.5; DB 4; Length 1601;
Best Local Similarity 17.6%; Pred. No. 2.7e-05;
Matches 223; Conservative 157; Mismatches 416; Indels 471;

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Gen search, using sw model

April 22, 2004, 12:06:19 ; Search time 64 Seconds
(without alignments)
5408.136 Million cell updates/sec

JS-09-501-171A-4

MFARXPGAGPLGAMPVDPQ.....PYSELNYETSHYPASPSWSV 1225

3LOSUM62

Gapop 10.0 , Gapext 0.5

1586107 seqs, 282547505 residues

its satisfying chosen parameters: 1586107

angth: 0

angth: 2000000000

Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

A_Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

is the number of results predicted by chance to have a
score than or equal to the score of the result being printed,
divided by analysis of the total score distribution.

SUMMARIES

% Query Match	Length	DB	ID	Description
100.0	1225	3	AAB07973	AAB07973 A human n
100.0	1233	4	ABG04996	ABG04996 Novel hum
96.1	1247	3	AAB07974	AAB07974 A murine
88.3	1084	2	AAY23900	AAY23900 Human res
85.1	1040	2	AAY24559	AAY24559 Presenili
61.4	756	2	AAG60664	AAG60664 Human ALA
44.2	1192	2	AAY23899	AAY23899 Human res
44.2	1211	2	AAY24560	AAY24560 Presenili
44.2	1211	3	AAY92336	AAY92336 Human p00
44.2	1211	5	ABE79001	ABE79001 Human p00
41.8	1110	4	AAE03648	AAE03648 Human ext
36.0	785	5	ABP41836	ABP41836 Human ova
35.9	696	4	ABG04995	ABG04995 Novel hum
26.7	1009	4	ABG20820	ABG20820 Novel hum
23.4	537	7	ADD27712	ADD27712 Human adi
22.4	785	4	ABG20821	ABG20821 Novel hum
22.2	962	6	ADA27163	ADA27163 ARVCF pro
21.5	939	4	AAE10797	AAE10797 Human cat
21.4	941	7	ADB75256	ADB75256 Prostate
20.1	834	6	ADA54659	ADA54659 Human pro
18.4	779	4	ABE66318	ABE66318 Drosophi
13.1	837	4	AAU27825	AAU27825 Human ful
13.1	837	3	ABG36464	ABG36464 Human pla
12.4	472	6	ABU70901	ABU70901 Human adi
12.2	821	3	ABG36467	ABG36467 Xenopus l

26	770.5	12.1	797	3	AAB36462	AAB36462
27	770.5	12.1	797	4	AAM79226	AAM79226
28	770.5	12.1	797	6	ABU56576	ABU56576
29	766	12.0	797	3	AAB36466	AAB36466
30	719.5	11.3	726	3	AAB36463	AAB36463
31	719.5	11.3	726	6	ABU56444	ABU56444
32	702	11.0	817	4	AAM80210	AAM80210
33	684	10.7	425	3	AAAB43829	AAAB43829
34	581.5	9.1	444	4	ABG18885	ABG18885
35	475	7.4	86	4	AAM19695	AAM19695
36	475	7.4	86	4	ABB39418	ABB39418
37	475	7.4	86	4	ABB32941	ABB32941
38	475	7.4	86	4	ABB24194	ABB24194
39	475	7.4	86	4	AAM72710	AAM72710
40	475	7.4	86	4	AAM60099	AAM60099
41	475	7.4	86	4	ABG54408	ABG54408
42	475	7.4	86	5	ABG42534	ABG42534
43	452	7.1	198	6	ADA54263	ADA54263
44	407.5	6.4	1266	4	ABG10319	ABG10319
45	396	6.2	193	4	ABG20816	ABG20816

ALIGNMENTS

RESULT 1

AAB07973

ID AAB07973 standard; protein; 1225 AA.

XX AC AAB07973;

XX DT 14-NOV-2000 (first entry)

XX DE A human neural plakophilin related armidillo protein.

XX KW Human; presenilin; PS1; PS2; nerve cell growth; nerve damage; h
neural plakophilin related armidillo protein; Alzheimer's disea
Parkinson's disease; amyotrophic lateral sclerosis; ALS; stroke
multiple sclerosis; ischemia; stroke; neural paropathy; sciatic
motor neuron disease; peripheral neuropathy; neuropathy; diabet
spinal cord injury; facial nerve crush.

XX OS Homo sapiens.

XX PN WO200047615-A2.

XX PD 17-AUG-2000.

XX PF 11-FEB-2000; 2000WO-CA000126.

XX PR 12-FEB-1999; 99US-0119835P.

XX PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.

XX PI St George- Hyslop PH, Fraser PE;

XX DR WPI; 2000-524531/47.

XX PS N-PSDB; AAA59700.

XX PT Stimulation of nerve cell growth using human Neural Plakophilin
Armidillo Protein (hNPRAP) polypeptide, useful for the treatment
diseases such as Alzheimer's, Parkinson's, and stroke.

XX PS Claim 3; Page 24-27; 33pp; English.

XX CC The present sequence represents human Neural Plakophilin Relate
Armidillo Protein (hNPRAP) polypeptide. hNPRAP interacts with p
(PS) I and II (PS1 and PS2). The specification describes a meth
stimulating the growth of nerve cell, comprising contacting th
hNPRAP. The hNPRAP polypeptide and polynucleotide are useful fo
nerve damage caused by a variety of diseases or physical trauma
including Alzheimer's disease, Parkinson's disease, amyotrophic
sclerosis (ALS), multiple sclerosis, stroke, ischemia associate

[illegible]

forensics, gene mapping, identification of mutations or genetic disorders or other traits to assess biodiversity or other types of data and products dependent on DNA and sequences. AB00010-AB030377 represent novel human diagnostic sequences of the invention. Note: the sequence data for this t appear in the printed specification, but was obtained in rmat directly from WIPO at pub/published_pct_sequences

AA;

100.0%; Score 6377; DB 4; Length 1233;
arity 100.0%; Pred. No. 0;
onservative 0; Mismatches 0; Indels 0; Gaps 0;

KPPGAAPLGAMPVDPQSSASEKSSLSPLGNTSNGDSETEETTSAILASVKEQEL 60
KPPGAAPLGAMPVDPQSSASEKSSLSPLGNTSNGDSETEETTSAILASVKEQEL 68
LTRLEAERQIVASQLERCKLGSETGSMSSMSAEQFQWOSQDQKDIIEDELTTG 120
LTRLEAERQIVASQLERCKLGSETGSMSSMSAEQFQWOSQDQKDIIEDELTTG 128
DSCIRSQESGILDPQDYSTGERPSLSQSALQLNSKPEGSFOYPASVHSNQTAL 180
DSCIRSQESGILDPQDYSTGERPSLSQSALQLNSKPEGSFOYPASVHSNQTAL 188
PSQLPARGTQARATQOSQCTTSRACHLAGPEPAPPPPPPPPPPPPPPPPPPP 240
PSQLPARGTQARATQOSQCTTSRACHLAGPEPAPPPPPPPPPPPPPPPPPPP 248
PAAAAAALYYSSSTLPAPRGGSPLAAPQGGSPTKLQGGSAPEGATVAAPRGSSP 300
PAAAAAALYYSSSTLPAPRGGSPLAAPQGGSPTKLQGGSAPEGATVAAPRGSSP 308
SRLAKSYSTSPINIVSSAGLSPIRVTSPTVQSTISSPIHQLSSITGYATLS 360
SRLAKSYSTSPINIVSSAGLSPIRVTSPTVQSTISSPIHQLSSITGYATLS 368
LVHASEQYKHSQELVATATLQPGSLAAGSRASVSSQHGHLGPELRALQSPHHI 420
LVHASEQYKHSQELVATATLQPGSLAAGSRASVSSQHGHLGPELRALQSPHHI 428
EDRVYQKPMRSLQSQDPLPPTAHTGYRTSTAPSSPGVDSVPLQRTSQHGPN 480
EDRVYQKPMRSLQSQDPLPPTAHTGYRTSTAPSSPGVDSVPLQRTSQHGPN 488
TFQRASYAAGPASNYADPYRQIQCPSVESPYSKSGPALPPEGTILARSFDSIQK 540
TFQRASYAAGPASNYADPYRQIQCPSVESPYSKSGPALPPEGTILARSFDSIQK 548
FGWRDPELPEVITOMLOHOPSPVQSNAAAYLOHLCPGDNKIKAELRROGGIOLLVDL 600
FGWRDPELPEVITOMLOHOPSPVQSNAAAYLOHLCPGDNKIKAELRROGGIOLLVDL 608
MTEVHRSACGALNLVYGKANDDNKIALKNCGGIPALVRLLRKTTDLRELVTGV 660
MTEVHRSACGALNLVYGKANDDNKIALKNCGGIPALVRLLRKTTDLRELVTGV 668
SSCDLKMPTIOTDALAVTNAVITPHSGWENSPLQDDRKIQLHSSQVLRNATGCLR 720
SSCDLKMPTIOTDALAVTNAVITPHSGWENSPLQDDRKIQLHSSQVLRNATGCLR 728
IAGEEARRRMRCDGTLTALLVYIQSALGSSEIDSKTNCVCLNLSVRLAAETS 780
IAGEEARRRMRCDGTLTALLVYIQSALGSSEIDSKTNCVCLNLSVRLAAETS 788
IMGTDELGLLCCGANGKDAESSGCWKKKKKKKQDQDGVGFLPDCAPPKGIOM 840
IMGTDELGLLCCGANGKDAESSGCWKKKKKKKQDQDGVGFLPDCAPPKGIOM 848
STIVKPYLITLLSECSNPDITLGAAGALQNLAAAGSWKSVYIRAAVVRKXKGLPILVEL 900

Db 849 LWHFSIVKPYLITLLSECSNPDITLGAAGALQNLAAAGSWKSVYIRAAVVRKXKGLP;
Qy 901 LRINDRVVCAVATARNMALDVNRKELIGYAMRDLVHRLPGGNNNSNTASKAM;
Db 909 LRINDRVVCAVATARNMALDVNRKELIGYAMRDLVHRLPGGNNNSNTASKAM;
Qy 961 TAVCCTTLEHVIITKMNENAKALRDAGGIEKLVGISKSGDKHSPKVVKAAASQVLNS;
Db 969 TAVCCTTLEHVIITKMNENAKALRDAGGIEKLVGISKSGDKHSPKVVKAAASQVLNS;
Qy 1021 DLRSLYKDGWSQYHFVASSSTIERDRQRPYSSSRTSPSISPVRSPPNRSASAPAK;
Db 1029 DLRSLYKDGWSQYHFVASSSTIERDRQRPYSSSRTSPSISPVRSPPNRSASAPAK;
Qy 1081 ISLKERKTDYECTGSNATYHGAKGEHTSRKDAMTAQNTGISTLYNSYGAPEADII;
Db 1089 ISLKERKTDYECTGSNATYHGAKGEHTSRKDAMTAQNTGISTLYNSYGAPEADII;
Qy 1141 SAQVPQPSRKDYETYPQFQNSTRNVDSPFQVHHRPPASEYTMHLGLKSTGI;
Db 1149 SAQVPQPSRKDYETYPQFQNSTRNVDSPFQVHHRPPASEYTMHLGLKSTGI;
Qy 1201 YSAARPYSELNYETSHYEPASDPSWV 1225
Db 1209 YSAARPYSELNYETSHYEPASDPSWV 1233

RESULT 3

AAB07974

ID AAB07974 standard; protein; 1247 AA.

XX AAB07974;

XX AC AAB07974;

XX DT 14-NOV-2000 (first entry)

XX DE A murine neural plakophilin related armidillo protein.

XX Human; presenilin; PS1; PS2; nerve cell growth; nerve damage; hN1
XX neural plakophilin related armidillo protein; Alzheimer's disease;
XX Parkinson's disease; amyotrophic lateral sclerosis; ALS; stroke;
XX multiple sclerosis; ischemia; stroke; neural paropathy; sciatic;
XX motor neuron disease; peripheral neuropathy; neuropathy; diabetic;
XX spinal cord injury; facial nerve crush.

XX Mus sp.

XX OS WO2000047615-A2.

XX PN 17-AUG-2000.

XX PD 11-FEB-2000; 2000WO-CA000126.

XX PF 12-FEB-1999; 99US-0119835P.

XX PR (UTOR) UNIV TORONTO GOVERNING COUNCIL.

XX PA St George- Hyslop PH, Fraser PE;

XX PI WPI: 2000-524531/47.

XX DR N-PSDB; AAA59701.

XX PT Stimulation of nerve cell growth using human Neural Plakophilin I

XX PT Armidillo protein (hNPRAP) polypeptide, useful for the treatment

XX PT diseases such as Alzheimer's, Parkinson's, and stroke.

XX PS Disclosure; Page 30-33; 33pp; English.

XX CC The present sequence represents Neural Plakophilin Related Armid-

XX CC Protein (NPRAP) polypeptide. Human NPRAP interacts with presenil-

XX CC I and II (PS1 and PS2). The specification describes a method for

XX CC stimulating the growth of nerve cells, comprising contacting ther

XX CC hNPRAP. The hNPRAP polypeptide and polynucleotide are useful for

XX CC nerve damage caused by a variety of diseases or physical traumas.

zheimer's disease, Parkinson's disease, amyotrophic lateral
 (s), multiple sclerosis, stroke, ischemia associated with
 al parapathy, motor neuron diseases, sciatic crush,
 neuropathy, neuropathy associated with diabetes, spinal cord
 facial nerve crush

7 AA;

96.1%; Score 6129; DB 3; Length 1247;

larity 94.8%; Pred. No. 0;

Conservative 8; Mismatches 29; Indels 28; Gaps 3;

RKPGGAALGAMPVDPQSSASEKTSLSPLNTSNGDGETETTSAILASVKEQEL 60

RKQSGAALGAMPVDPQSSASEKNSLSPLNTSNGDGETETTSAILASVKEQEL 60

RLTRELEAERQIVASQLERCKLGSETGSMSSMAEBOFQWQSQDQKQDELTG 120

RLTRELEAERQIVASQLERCKLGSETGSMSSMAEBOFQWQSQDQKQDELTG 120

VDSCIRLSQESGILDPODYSTGERPSLLSQALQNSKPEGSFOYPASVHNSQTLAL 180

VDSCIRLSQESGILDPODYSTGERPSLLSQALQNSKPEGSFOYPASVHNSQTLAL 180

TPSQLPARQTARATGQSFSCQTTSRAGHLAGEPAPPPPPPPPPPPPPPSLGSAPHL 240

APSQLPARSTQARAAGQSFSCQTTGRAGHLAGEPA-PPPPPPPPPPPPPPPSLGSAPHL 239

APAAAALYSSSTLPPPPGSGPLAAPQGGSPKLOGGSAPGATYAAPRGSSP 300

APP--AAAALYSSSTLPPPPGSGPLTTTQGGSPKLOGGSAPGATYAAPRGSSP 297

PSRLAKSYSTSPINIVSSAGLSPIRVTSPPTVQSTISSPIHOLSTIGTYATLS 360

PSRLAKSYSTSPINIVSSAGLSPIRVTSPPTVQSTISSPIHOLSTIGTYATLS 357

RLVHASEQYKHSQELYTATLQRFGLAAGSRASYSSQHGHLGPELALQSPHEHI 420

RLVHASEQYKHSQELYTATLQRFGLAAGSRASYSSQHGHLGPELALQSPHEHI 417

YEDRVYQPPMPSLSQSDPLPPAHTGYRTSTAPSGVDSPVLPQTGSHGQPN 480

YEDRVYQPPMPSLSQSDPLPPAHTGYRTSTAPSGVDSPVLPQTGSHGQPN 477

YATFORASYAAGPASYADPYRQLYCPSYESPSYSGSPALPPEGLTARSPTSISQK 540

YATFORASYAAGPASYADPYRQLYCPSYESPSYSGSPALPPEGLTARSPTSISQK 537

REFGWRDPPELPEVIQMLQHPFVSQSNAAAYLQHLGFGDNKIKAEIRROGGIQLVYL 600

REFGWRDPPELPEVIQMLQHPFVSQSNAAAYLQHLGFGDNKIKAEIRROGGIQLVYL 597

IRMTVHRSGAGLRNLVYGKANDDNKIALKNCGGIPALVRLRKTDLREILVTGV 660

IRMTVHRSGAGLRNLVYGKANDDNKIALKNCGGIPALVRLRKTDLREILVTGV 657

YJSSCDALXMPITQDALAVLTNAVITPHSGWENSPLODDRKIQHSSQVLRNATGCLR 720

YJSSCDALXMPITQDALAVLTNAVITPHSGWENSPLODDRKIQHSSQVLRNATGCLR 717

SSAGEARRMRSCDGLTDALLVYIOSALGSSRIDSKVTENCVCILRNLSYRLAAETS 780

SSAGEARRMRSCDGLTDALLVYIOSALGSSRIDSKVTENCVCILRNLSYRLAAETS 777

YHMGTDLDGLGCEANGDAESSGCKKKKKKQDQWDGVGVLPCDAEPPKGIQM 840

YHMGTDLDGLGCEANGDAESSGCKKKKKKQDQWDGVGVLPCDAEPPKGIQM 837

YPSIVKPYLTLSECSNPDTLEGAAGALQNLAAAGSWK----- 879

YPSIVKPYLTLSECSNPDTLEGAAGALQNLAAAGSWK----- 897

--WSVYIRAAVRKEKGLPILVELLRIDNDRVCAVATALRNALDVRNKLICKYMR 935

898 CLPQWSVYIRAAVRKEKGLPILVELLRIDNDRVCAVATALRNALDVRNKLICKYMR 935

936 DLVHRLPGCGNSNNTASKAMSDDTVTAVCCITLHEVITKMNENAKALRDAGGIEKI 958

958 DLVHRLPGCGNSNNTASKAMSDDTVTAVCCITLHEVITKMNENAKALRDAGGIEKI 996

996 SKGDKHSPKVVVKAASQVLNSMWQYRDLRSLYKKDQWGSQYHFVASSSTIERDRQRI 1018

1018 SKGDKHSPKVVVKAASQVLNSMWQYRDLRSLYKKDQWGSQYHFVASSSTIERDRQRI 1056

1056 TPSISPVVRVSPNNRSASAPASPREMISLKERKTDYECTGSGNATYHGAKGHTSRI 1078

1078 TPSISPVVRVSPNNRSASAPASPREMISLKERKTDYECTGSGNATYHGAKGHTSRI 1116

1116 QNTGISTLYNSVGCAPAEEDIKHQVSAQPVQPEPSRKDYETQPFQFONSTRYDEK 1138

1138 QNTGISTLYNSVGCAPAEEDIKHQVSAQPVQPEPSRKDYETQPFQFONSTRYDEK 1176

1176 VHRPPASEYTMHLGLKSTGNYVDFYSAARPYSSELNYTSHYPASPDWV 122:

122: VHRPPASEYTMHLGLKSTGNYVDFYSAARPYSSELNYTSHYPASPDWV 124:

RESULT 4

AAV23900

ID AAV23900 standard; protein; 1084 AA.

XX AC AAV23900;

XX DT 27-SEP-1999 (first entry)

XX DE Human resenilin binding armadillo protein GT24/hnPRAP.

XX KW Human; presenilin 1; P51; presenilin-binding protein; interacti

XX KW presenilin allele; Alzheimer's disease; senile dementia;

XX KW psychiatric disease; schizophrenia; depression; neurological di

XX KW stroke; cerebral haemorrhage; p0071; armadillo protein.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Misc-difference 115 /note= "not specified"

XX FT WO9935501-A1.

XX PD 15-JUL-1999.

XX PF 08-JAN-1999; 99WO-CA000018.

XX PR 09-JAN-1998; 98US-0070948P.

XX (UTOR) UNIV TORONTO GOVERNING COUNCIL.

XX PI St George- Hyslop PH, Fraser PE;

XX DR WPI; 1999-419410/35.

XX PT Identifying substances that alter presenilin interactions, usef

XX PT screening individuals for presenilin alleles associated with Al

XX PT disease - useful for diagnosis of Alzheimer's disease.

XX PS Claim 9; Fig 1; 40pp; English.

XX The present sequence represents a human presenilin-binding protei

XX termed GT24/human neuronal plakoglobin related armadillo protei

XX hnPRAP). The specification describes a method for identifying s

XX that alter the interaction of a presenilin with a presenilin-bi

XX protein. The method comprises contacting the interacting domain

XX presenilin protein to a presenilin-binding protein in the prese

XX test substance, and measuring the interaction of the presenilin

XX presenilin-binding protein. The method can be used to screen in

n alleles associated with Alzheimer's disease and related
ch as senile dementia's, psychiatric diseases such as
and depression, and neurological disease, such as stroke
haemorrhage

AA;

88.3%; Score 5633; DB 2; Length 1084;

arity 99.8%; Pred. No. 0;

conservative 0; Mismatches 2; Indels 0; Gaps 0;

SQSALQNSKPEGSFQVPSYHSNQTALGETTSPQLPARCTQARATGQSFSQGT 207
SQSALQNSKPEGSFQVPSYHSNQTALGETTSPQLPARCTQARATGQSFSQGT 66
HLAGEPAPPP 267
HLAGEPAPPP 126
PQGSPTKLQSGSAPEGATYAAPRGSSPKQSPSLAKSYSTSSPINIVSSAGLS 327
PQGSPTKLQSGSAPEGATYAAPRGSSPKQSPSLAKSYSTSSPINIVSSAGLS 186
TSPTVQSTISSPPIHLSSTIGVATLSPTKRLVHASEQYKSHQELIYATLQR 387
TSPTVQSTISSPPIHLSSTIGVATLSPTKRLVHASEQYKSHQELIYATLQR 246
AAGRASYSOHHGLPELALQPEHHIDPIYEDRVYQKPPMRSLQSQGDPLPP 447
AAGRASYSOHHGLPELALQPEHHIDPIYEDRVYQKPPMRSLQSQGDPLPP 306
TYRTSTAPSPGVDSVPLQRTSGHGPQNAATAATQFASAAAGPASNYADPYRLQ 507
TYRTSTAPSPGVDSVPLQRTSGHGPQNAATAATQFASAAAGPASNYADPYRLQ 366
VESPYKSGPALPPEGTIARSPIDSIQKDPREFGWRDPPELPIVQLMHOFPSPVQ 567
VESPYKSGPALPPEGTIARSPIDSIQKDPREFGWRDPPELPIVQLMHOFPSPVQ 426
AYLOHLCFGDNKIKAEIRROGGIQLLVLLDHRMTEVHRSACGALRNLYVGKAND 627
AYLOHLCFGDNKIKAEIRROGGIQLLVLLDHRMTEVHRSACGALRNLYVGKAND 486
LKNCGGIPALVRLKRTDLREIRELVGTGLVNLSSCDALXMPPIQDALAVLTNAV 687
LKNCGGIPALVRLKRTDLREIRELVGTGLVNLSSCDALXMPPIQDALAVLTNAV 546
IGWENSPLODDRKIQLHSSQVLRNATGCLRNVSAGEARRMRRECGLTDALLYVI 747
IGWENSPLODDRKIQLHSSQVLRNATGCLRNVSAGEARRMRRECGLTDALLYVI 606
IGSSEIDSKTVENCVCILRNLSYRLAAETSGOQMGCTDELGLLCEANGKDAESSG 807
IGSSEIDSKTVENCVCILRNLSYRLAAETSGOQMGCTDELGLLCEANGKDAESSG 666
JXKKKKKSQDQWGVGFLPDCAPPPKGIQMLMHPISIVKPYLTLLSECSNPTLEGAA 867
JXKKKKKSQDQWGVGFLPDCAPPPKGIQMLMHPISIVKPYLTLLSECSNPTLEGAA 726
JNLAAGSWKWSVYIRAAVRKEGLPLVLELLRINDNRVVCVATARNALDVRNKE 927
JNLAAGSWKWSVYIRAAVRKEGLPLVLELLRINDNRVVCVATARNALDVRNKE 786
CYAMRDLVHRLPGNNNSNTASKAMSDDTVTAVCCITLHEVITKMNNAKALRDAIGI 987
CYAMRDLVHRLPGNNNSNTASKAMSDDTVTAVCCITLHEVITKMNNAKALRDAIGI 846
JGISKSGDKRHSKPVKAAASQVLNSMWQYRDLRSYKKGWSQYHFVASSSTIERDR 1047
JGISKSGDKRHSKPVKAAASQVLNSMWQYRDLRSYKKGWSQYHFVASSSTIERDR 906
JSSSRTPSPVVRVSPNNRSASAPSPREMI SLKERTDYECTGSGNATHYHGGE 1107

Db 907 QRPVSSRTSPISPVVRVSPNNRSASAPSPREMI SLKERTDYECTGSGNATHYHGGE
Qy 1108 SRKDAMTAQNTIGISTLYRNSYGAPAEIDIKHNQVSAQPVQPBPSPKDYETQYQFQNI
Db 967 SRKDAMTAQNTIGISTLYRNSYGAPAEIDIKHNQVSAQPVQPBPSPKDYETQYQFQNI
Qy 1168 DESFFEDQVHRRPPASEYTMHLGLKSTGNYVDYFSAARPPYSELNYETSHYPASPD
Db 1027 DESFFEDQVHRRPPASEYTMHLGLKSTGNYVDYFSAARPPYSELNYETSHYPASPD
RESULT 5
AAW24559
ID AAW24559 standard; protein; 1040 AA.
XX AAW24559;
AC AAW24559;
XX
DT 06-FEB-1998 (first entry)
XX
DE Presenilin-interacting protein GT24.
XX
KW Presenilin-interacting protein; human; Alzheimer's disease; diag
therapy; transgenic animal; animal model; GT24.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 12 /note= "encoded by GST"
FT Misc-difference 40 /note= "encoded by CCR"
FT Misc-difference 71 /note= "encoded by AKC"
FT Domain 346..862 /note= "presenilin-interacting domain"
XX WO9727296-A1.
XX 31-JUL-1997.
XX 27-JAN-1997; 97WO-CA000051.
XX 26-JAN-1996; 96US-00592541.
PR 05-JUL-1996; 96US-0021673P.
PR 12-JUL-1996; 96US-0021700P.
PR 08-NOV-1996; 96US-0029895P.
PR 02-JAN-1997; 97US-0034590P.
XX (HSCR-) HSC RES & DEV LP.
PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.
XX
PI St George- Hyslop PH, Fraser PE, Rommens JM;
XX WPI: 1997-393684/36.
DR N-PSDB; AAT79967.
XX
PT Presenilin-interacting protein genes - used to develop products
diagnosis, therapy and study of Alzheimer's disease and related
disorders.
PS Claim 1; Page 99-101; 133pp; English.
XX
CC GT24, a human protein with 'armadillo' repeats, has been identif
presenilin-interacting protein (PIP). A yeast two-hybrid kit was
to screen a human brain cDNA library for clones which interacted
presenilin transmembrane 6-7 loop domain; mutations in this loop
are known to be causative of Alzheimer's disease (AD). 9 PIP gen
sequences (see AAT79966-74) including GT24 were identified. PIP
acids, PIP proteins and peptides (especially the presenilin inte
domain), antibodies to PIPs, cells transformed with PIP nucleic
and transgenic animals altered with PIP nucleic acids can be use
diagnosis, therapy and study of AD and related disorders. They c
used to identify compounds which can modulate the expression of

sh bind to a PIP or modulate its activity

10 AA;

85.1%; Score 5426; DB 2; Length 1040;

larity 99.6%; Pred. No. 0;

Conservative 0; Mismatches 4; Indels 0; Gaps 0;

.PARGTQARATGSGFSGTTSRAGHLAGPPAPPPPPPPPPPPPPPSLGSFAFHLDPDAPP 245

.PARGTQARATGSGFSGTTSRAGHLAGPPAPPPPPPPPPPPPPPSLGSFAFHLDPDAPP 60

AAALYSSSTLPAPRGGSPPLAAGPGSTKLQSGSAPGATYAAPRGSSPKQSPS 305

AAALYSSSTLPAPRGGSPPLAAGPGSTKLQSGSAPGATYAAPRGSSPKQSPS 120

AKSYSTSPINIVSSAGLSPIRVTSPTVQSTISSPIHQLSSTIGTVATLSPTKRL 365

AKSYSTSPINIVSSAGLSPIRVTSPTVQSTISSPIHQLSSTIGTVATLSPTKRL 180

ASEQYKSHQELVATATLQPGSLAGSRASYSSQGHGHPALQSPPEHHIDPIYE 425

ASEQYKSHQELVATATLQPGSLAGSRASYSSQGHGHPALQSPPEHHIDPIYE 240

AYOKPMRSUSQSGDPLPAHTGTVRTSTAPSSPGVDSVPLQRTGSHQGPQNAAT 485

AYOKPMRSUSQSGDPLPAHTGTVRTSTAPSSPGVDSVPLQRTGSHQGPQNAAT 300

ASAYAGPASNYADPYRQLQYCFVSFSPYSGKGPALPPEGTARSIDSIGKDPREF 360

ASAYAGPASNYADPYRQLQYCFVSFSPYSGKGPALPPEGTARSIDSIGKDPREF 545

ADPELVEIOMLOHQPSPVOSNAAYLQHLCFGDNKIKABIRQGGIQLLVLDLDRM 605

ADPELVEIOMLOHQPSPVOSNAAYLQHLCFGDNKIKABIRQGGIQLLVLDLDRM 420

JHSACGALRNLYVGRANDDNKIALKNCGGIPALVRLRKTDTLEIRLVTVGLWNLS 665

JHSACGALRNLYVGRANDDNKIALKNCGGIPALVRLRKTDTLEIRLVTVGLWNLS 480

DALKMPIIDALAVLTNAVITPHSGWENSPLODDRKIQLHSSQVLRNATGCLRNSSA 725

DALKMPIIDALAVLTNAVITPHSGWENSPLODDRKIQLHSSQVLRNATGCLRNSSA 540

BARRRRECGLTDALLYVQTSALGSSEIDSKTVENCVCILRNLSYRLAETSGQGHM 785

BARRRRECGLTDALLYVQTSALGSSEIDSKTVENCVCILRNLSYRLAETSGQGHM 600

DELDDLGCGRANGKDAESSGCWKKKKKSKQDQWGVPLPDCAPPKGIQMLWHP 845

DELDDLGCGRANGKDAESSGCWKKKKKSKQDQWGVPLPDCAPPKGIQMLWHP 660

KPYLTLLSECSNPDTLEGAGALQNLAAAGSKWSVYIRAAVRKEKGLPIVLLELRIDN 905

KPYLTLLSECSNPDTLEGAGALQNLAAAGSKWSVYIRAAVRKEKGLPIVLLELRIDN 720

VCAVATARNMALDVNRKELIGYAMRDVLVHRLPGGNNNTASKAMSDDTVTAVCC 965

VCAVATARNMALDVNRKELIGYAMRDVLVHRLPGGNNNTASKAMSDDTVTAVCC 780

HEVITKNENAKALRDAGGIEKLVGISKSGDKHSPKVVKAASQVLNSMWMQYDLRSL 1025

HEVITKNENAKALRDAGGIEKLVGISKSGDKHSPKVVKAASQVLNSMWMQYDLRSL 840

KDQWSQYHFAVSSSTIERDRQRPVSSRTSPISFVRVSPNRRSASAPASPREMISLKE 1085

KDQWSQYHFAVSSSTIERDRQRPVSSRTSPISFVRVSPNRRSASAPASPREMISLKE 900

IDVECTGSNATYHGKGEHTSRKDMATQNTIGISTLYRNSYGAPAEIDKKNQVSAQPV 1145

IDVECTGSNATYHGKGEHTSRKDMATQNTIGISTLYRNSYGAPAEIDKKNQVSAQPV 960

SPSRKDYETQPFQNSRNYDESFEDQVHRRPPASEYTMHLGLKSTGNTVDFYSAAR 1205

Db 961 PQEFSRKYDYTFQFQNSRNYDESFEDQVHRRPPASEYTMHLGLKSTGNTV

QY 1206 PYSELNYETSHYPASPDWSV 1225

Db 1021 PYSELNYETSHYPASPDWSV 1040

RESULT 6

AAW60664

ID AAW60664 standard; protein; 756 AA.

XX AAW60664;

AC AAW60664;

DT 26-OCT-1998 (first entry)

XX Human ALARM protein.

XX ALARM; adherens-junction linked arm protein; delta-catenin; pre

diagnosis; Alzheimer's disease; antibody; human.

XX Homo sapiens.

Key Location/Qualifiers

FT Region 159..202

FT Region 203..243

FT Region 412..450

FT Region 459..499

XX WO9825142-A1.

XX 11-JUN-1998.

XX 02-DEC-1997; 97WO-US022093.

XX 02-DEC-1996; 96US-0031556P.

XX (BGM) BRIGHAM & WOMENS HOSPITAL.

XX Kosik KS, Zhou J;

XX WPI; 1998-333494/29.

XX N-PSDB; AAV38305.

XX Human ALARM polypeptide(s) binding to presenilin 1 - useful, e.

XX detect presenilin 1 and diagnose human diseases caused by mutant

XX protein interacting with ALARM.

XX Claim 4; Fig 1A-F; 62pp; English.

XX This novel human protein, named ALARM or delta-catenin, is able
to presenilin 1. Its amino acid sequence was deduced from a cDN.
(see AAV38305) isolated from a human brain cDNA library. ALARM
adherens-junction linked arm protein) is expressed almost exclu
brain tissue. It contains 4 copies of the arm repeat that is
characteristic of catenin proteins. Overall, ALARM and delta-ca
52.3% similar and 32.1% identical. The invention provides iso
polypeptide, nucleic acids, vectors, host cells and specific an
Presenilin 1 protein can be detected in samples (especially hum
cerebrospinal fluid), by contacting with ALARM (claimed). Mutat
presenilin 1 gene is associated with familial Alzheimer's disea
can be used to diagnose human diseases caused by a mutant form
protein interacting with ALARM, by analysing fluid samples to d
ALARM-interacting protein (claimed). ALARM can also be used to
altered levels of presenilin 1, by contacting a sample with ALA
determining bonding (claimed). ALARM polypeptides and antibody
used to detect an ALARM-containing complex in biological sample
contacting with polypeptide or antibody and determining whether
polypeptide/antibody binds to a sample component (claimed). The

e also useful to detect ALARM polypeptides and to measure candidate compounds on expression or localisation of ALARM. ul to evaluate engineered cells prior to introduction in to inhibit abnormal ALARM activity or to generate antibodies, useful therapeutically

AA;
61.4%; Score 3916; DB 2; Length 756;
arity 94.9%; Pred. No. 3.9e-271; Indels 36; Gaps 1;
conservative 0; Mismatches 4;
SOSQDPLPPAHTGYRTSTAPSPGVDSPVLPQTGSGHQPQNAATAATFORASYAA 493
SOSQGVLPPLAHTGYRTSTAPSPGVDSPVLPQTGSGHQPQNAATAATFORASYAA 60
INYADYRQLOPCPSVESYKSGPALPEGLTLARSPSIDSTQKOPREGWRDPPLP 553
INYADYRQLOPCPSVESYKSGPALPEGLTLARSPSIDSTQKOPREGWRDPPLP 120
MLQHOPFVSQNAAYLQHLCEGDNKIKAERQGGIQLLVLDLDRHTEVHRSA 613
MLQHOPFVSQNAAYLQHLCEGDNKIKAERQGGIQLLVLDLDRHTEVHRSA 180
NLVYGKANDDNKIALKNCGGIPALVRLRKTDLLEIRLVGTGLVNLSSCDALKMP 673
NLVYGKANDDNKIALKNCGGIPALVRLRKTDLLEIRLVGTGLVNLSSCDALKMP 240
ALAVLTNAVIIIPHSGWENSPLODRKIQLHSSQVLRNATGCLRNYSAGEARRM 733
ALAVLTNAVIIIPHSGWENSPLODRKIQLHSSQVLRNATGCLRNYSAGEARRM 300
GLTDLALLVIOGALGSSEIDSTVENCVCILRNLSYRLAAETSGOHHMTDDELGL 793
GLTDLALLVIOGALGSSEIDSTVENCVCILRNLSYRLAAETSGOHHMTDDELGL 360
ANGKDAESGCGWKKKKKKKQDQWDGVGPLPDCAEPKGIQMLWHPISVVKPYLT 853
ANGKDAESGCGWKKKKKKKQDQWDGVGPLPDCAEPKGIQMLWHPISVVKPYLT 420
SNPDTLEGAGALQNLAAAGSWKSVYIRAAVRKEKGLPLVLLRDLNDRVACAVA 913
SNPDTLEGAGALQNLAAAGSWKSVYIRAAVRKEKGLPLVLLRDLNDRVACAVA 480
RNALDVNRKELLGKYMARDLVRHLPGGNNNNNTASKAMSDDTVTAVCCTLHEVITK 973
RNALDVNRKELLGKYMARDLVRHLPGGNNNNNTASKAMSDDTVTAVCCTLHEVITK 540
YAKALRDAGGIEKLVGISKSGDKHPKVYKAAASQVLNSMWQYRDLRSLYKKGWSQ 1033
YAKALRDAGGIEKLVGISKSGDKHPKVYKAAASQVLNSMWQYRDLRSLYKKGWSQ 600
JASSSTIERDRQPYSSSRTPSISPVRSVNNNSASAPSPREMI SLKERTDYECT 1093
JASSSTIERDRQPYSSSRTPSISPVRSVNNNSASAPSPREMI SLKERTDYECT 660
ATHGKAGERTSRKDMATQNTGISTLYRNSYCAPAEDIKHQVSAQVPQEPSRKD 1153
ATHGKAGERTSRKDMATQNTGISTLYRNSYCAPAEDIKHQVSAQVPQEPSRKD 692
YQPFQNSTRYNDSFFEDQVHRPPASEYTMHLGKSTGNYVDYFYAARPYSEINYE 1213
-----NSTRYNDSFFEDQVHRPPASEYTMHLGKSTGNYVDYFYAARPYSEINYE 744
YPASPDWSW 1225
YPASPDWSW 756

ndard; protein; 1192 AA.

AC AAY23899;
XX 27-SEP-1999 (first entry)
XX Human resenilin binding armadillo protein p0071.
DE
XX
XX Human; presenilin 1; PS1; presenilin-binding protein; interactin
KW presenilin allele; Alzheimer's disease; senile dementia;
KW psychiatric disease; schizophrenia; depression; neurological dis
stroke; cerebral haemorrhage; p0071; armadillo protein.
XX Homo sapiens.
OS
XX WO9935501-A1.
PN
XX 15-JUL-1999.
PD
XX 08-JAN-1999; 99WO-CA0000018.
PF
XX 09-JAN-1998; 98US-0070948P.
PR
XX (UTOR) UNIV TORONTO GOVERNING COUNCIL.
PI
XX St George- Hyslop PH, Fraser PE;
XX WPI; 1999-419410/35.
XX
XX Identifying substances that alter presenilin interactions, usefu
screening individuals for presenilin alleles associated with Alz
disease - useful for diagnosis of Alzheimer's disease.
XX
XX Claim 10; Fig 1; 40pp; English.
XX
XX The present sequence represents a human presenilin-binding prote
termed p0071. It is a member of the armadillo family of proteins
specification describes a method for identifying substances that
the interaction of a presenilin with a presenilin-binding protei
method comprises contacting the interacting domain of a presenil
protein to a presenilin-binding protein in the presence of a tes
substance, and measuring the interaction of the presenilin and t
presenilin-binding protein. The method can be used to screen ind
for presenilin alleles associated with Alzheimer's disease and r
disorders, such as senile dementia's, psychiatric diseases such
schizophrenia and depression, and neurological disease, such as
and cerebral haemorrhage
XX
XX Sequence 1192 AA;
Query Match 44.2%; Score 2818.5; DB 2; Length 1192;
Best Local Similarity 49.2%; Pred. No. 3e-192;
Matches 618; Conservative 192; Mismatches 337; Indels 109;
QY 15 MPVDPQSSASEKTSLSPLNTSNGDSETETT-SAILASVKEQELQELRLTRE
Db 1 MPAEQASLVEGQPTQREA-ASTGPGNEPETTATILLASVKEQELQELRLTRE
QY 74 QIVASQLERCKLSETGSMSSSAEQQWQSQD-----QOKDIEDELTTGLE
Db 60 QIVASQLERCKLGAESPIASTSTSEKSPWRSTDVNTGSKPRVSDAVQPNNY
QY 128 IRSLSQSGIL-PPDYSTGERFSL--SQSALQASKPEGSFYPAHYHNTLA
Db 116 IRTEPEQGLYSPQOTSLHESGSLGNSRSTOMNSYSDSGYQAGSFNSQVNS
QY 180 -----LGETTPSQLPARGTQARATGQSFQ-GTTSRAGHLAGEPAPPPPP-----
Db 176 QQHSFICSTNNHV-----RNSRAEGGTLVQPSVANPAMERVSVSPSAQSPYVI
QY 225 PPREPPAPSLGSAFHLP---DAPPAALAAALYSSTLPPAPRGGSPLAAPQGG
Db 232 PSRGLSLTSLGSGFGSPVTDPRPLNPSA--YSSTTLPA-ARAASPY-SQRPAS
QY 282 RGSAPGATYAAPRGSSPK-QSPSLAKSYSTSSPINIVSSAGLSPIRVTSPF

andard; protein; 1211 AA.

interacting protein p0071.

XX	Presenilin-interacting protein; human; Alzheimer's disease; dia
KW	therapy; transgenic animal; animal model; p0071.
XX	OS
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
PH	509..1022
FT	/note= "presenilin-interacting domain"
FT	
XX	W09727296-A1.
PN	
XX	
PD	31-JUL-1997.
XX	
PF	27-JAN-1997; 97WO-CA000051.
XX	
PR	26-JAN-1996; 96US-00592541.
PR	05-JUL-1996; 96US-0021673P.
PR	12-JUL-1996; 96US-0021700P.
PR	08-NOV-1996; 96US-0029895P.
PR	02-JAN-1997; 97US-0034590P.
XX	
PA	(HSCR-) HSC RES & DEV LP.
PA	(UTOR) UNIV TORONTO GOVERNING COUNCIL.
XX	
PI	St George- Hyslop PH, Frazer PE, Rommens JM;
DR	WPI; 1997-393684/36.
DR	N-PSDB; AA179968.
XX	
PT	Presenilin-interacting protein genes - used to develop products
PT	diagnosis, therapy and study of Alzheimer's disease and related
PT	disorders.
XX	
XX	Claim 1; Page 104-107; 133pp; English.
XX	
CC	p0071, a human protein with 'armadillo' repeats, has been ident
CC	presenilin-interacting protein (PIP). A yeast two-hybrid kit wa
CC	to screen a human brain cDNA library for clones which interact
CC	presenilin transmembrane 6-7 loop domain; mutations in this loo
CC	are known to be causative of Alzheimer's disease (AD). 9 PIP ge
CC	sequences (see AA179966-74) including p0071 were identified. PI
CC	acids, PIP proteins and peptides (especially the presenilin int
CC	domain), antibodies to PIPs, cells transformed with PIP nucleic
CC	and transgenic animals altered with PIP nucleic acids can be us
CC	diagnosis, therapy and study of AD and related disorders. They
CC	used to identify compounds which can modulate the expression of
CC	gene or which bind to a PIP or modulate its activity
XX	
SQ	Sequence 1211 AA;
Query Match	44.2%; Score 2818.5; DB 2; Length 1211;
Best Local Similarity	49.2%; Pred. No. 3.1e-192;
Matches	618; Conservative 192; Mismatches 337; Indels 109;
Qy	15 MPVPDPSSASEKTSLSPLNTNSGDSGETTT-SAILASVKQELOFERLTR
Db	1 MPAEQASLVEEGQPOTROEA-ASTGPGMEPTATTILASVKQELOFQRLTR
Qy	74 QIVASQLERCKLGSSETGSMSSAAEEQFQMSQD-----GQKDIEDELRTGL
Db	60 QIVASQLERCLGAESPSTASTSTSEKSFPMWRTDVPNTGVSKRVSDAVQPNN
Qy	128 IRLSQESGIL-PPQYSYTGERSLLI--SQSALQNLNKEPGEFQYPASVHSNQTLL
Db	116 IRTEPQGTLYPEQTSLHESGSGLSNGRSSTOMNSYSDSGYQAGSFHNSQNV.
Qy	180 ----LGETTPSQLPARGTQARATQGSFSQ-GTTSRAGHLAGPPADPPPP----
Db	176 QQHSFTGSTNNHV-----RNSRAEQCTLVQPSVANRMRVSSVPSSAQSPSY
Qy	225 PPREPPAPSLGSAFHLP---DAPPAAAAAALYSGSSTLPAPPGRGSPPLAAPQGG.

SLRTSLGSGFGSPVTDPRPLNPSA---YSSITLPA-ARASPY-SQRPASPTAIR 286
 APEGATVAAPRGSGPK-QSPRLAKSXTSSPINIVSSAGLSPIRVTSPTVQST 340
 JTRQOT-SNPGPTPOQTARV-----GSPILTDAQT-----RVASPSQGO-V 333
 PIHQLSSTIGTYATLSPT-KRLVHASQYKSHSOELYATATLQRPGLSAGSRASY 399
 P--KRSQMTAVPQHLGFLQRTVHDMQFGQQQYDIYERMPVPPDLSL-TGLRSSY 390
 ZHIGPELRAIQSPHHIDPIYEDRVYQKPMRSLSQSGDPLPPAHTGTTTSTAP 459
 SOLGQDLRSVSPDLHTPTIEGRTYSPYRSPNHGTVL-LOGSQTALYRT---- 445
 VDSV-PLQRTGSHQGPQNAATAFORASYAAGPASNYADPVROLQYCPSPVYSK 518
 VSGIGNLQRTSSQR-----STLTQRYNYALNTTATYAEPIQYRVQ-ECYNR 496
 LPPEGTLARSIDSIDQDPREFQWRDPELPEVIQMLQHQFPSPVQSNAAAYLQHL 578
 VPADDGTRSPSIDSIDQDPREFQWRDPELPEVIHMLHQFPSPVQNAAYLQHL 556
 KIKABIRROGGIQLLVLLDHRMTEVRSAGCALRNLYGKANDDNKIALKNCGGI 638
 KVMVEVCLGGIKHLVLLDHRVLEVKQKAGCALRNLYGKSTDENKIAMKNVGGI 616
 RLLRKTDLREIRELVTVGLVNLSSCDALKMPIDQALAVLTNAVIIIPHSWENSPL 698
 RLLRKSIDAEREVLTVGLVNLSSCDALKMPIDQALAVLTNAVIIIPHSWENSPL 676
 KIQHSSQVLRNATGCLRNVSAGAEARRMRECDGLTDLALYVQISALSGSEIDS 758
 KIKFQTSILVRNTGCLRNLTSAEAEARKQMRSCGVLGVDLSLYVHCVTNTSDYDS 736
 NCVCILNLSVRLAAETISQOHMGTDELGLLCEANGKDAESGCGWKKKKKK- 817
 NCVCILNLSVRLAAETISQOHMGTDELGLLCEANGKDAESGCGWKKKKKK- 795
 DQWDGVLPLDCABPPKGIQMLHPSIVKPYLTLLSCSPDTEGAGALQNLAA 875
 DQWDGVLPLDCABPPKGIQMLHPSIVKPYLTLLSCSPDTEGAGALQNLAA 855
 WSVYIRAAVKEKGLPILVELLRIDNRVVCATARNALDVRNKEKLGKVMR 935
 FAAYIRGRPRKGLPILVELLRIDNRVVCATARNALDVRNKEKLGKVMR 915
 RLPGNNNSNTASKMSDDTVTAVCCTLHEVITKNMENAKALRDAGIEKLVGISK 995
 RLPGNGGPS-----VLSDETMAALCCALHEVITKNMENAKALDAGIEKLVNITK 970
 KHPKVVKAASQVLNMQYRDLRLSKYKQGSQYHFVASSSTTIERDRQRPYSRR 1055
 RSSLKVKAQAQVLTWLQYRDLRLSKYKQGSQYHFVASSSTTIERDRQRPYSRR 1026
 SPV--RVSPNNRSASAPSPREMTSLKERTDYCTGTSNATYHGAKEHTSRKDM 1113
 STTNQOMSPIIQSGVSTSSPALLGIRDPRSEYDRTQPMQYNSQGDATHK--GL 1083
 TGISTLYRNSYGAPED-----IKENQVSAQVPQEPSRKQVETYPQCNSTRNDE 1169
 SKPFIYIYSSPANNRRLQHQQLYYS--QDDSNRKNPDVRLYLQSPHSYED 1141
 DQVHRRPPASEYTHLGLKSTGNTVDFYSAARPYSELYETSHYPASPDSWV 1225
 DRVHF-PASTDYSTQYGLKSTNTYVDFYSTKRP---SYRAEQYPGSPDSWV 1192

dard; protein; 1211 AA.

21-AUG-2000 (first entry)
 Human p0071.
 NIK1 Interacting Protein; p0071; protein complex; cytostatic; ant
 neuroprotective; cardiant.
 Homo sapiens.
 W0200020448-A2.
 13-APR-2000.
 06-OCT-1999; 99WO-US023314.
 06-OCT-1998; 98US-00167206.
 (CURA-) CURAGEN CORP.
 Nandabalan K, Schulz VP, Yang M;
 WPI; 2000-303742/26.
 N-PSDB; AAA09308.
 New complex of a NIK1 protein and a NIK1 protein-interacting prot
 useful for treating cancer, hyperproliferative disorder,
 neurodegenerative disorder, cardiomyopathies, viral infections ar
 metabolic disorders.
 Example 1; Page 154-159; 172pp; English.
 AAY92331-37 were isolated in a modified yeast two hybrid system v
 NIK1 protein as "bait". These are known sequences which are NIK1
 interacting proteins. The invention concerns purified complexes c
 protein and a NIK1 protein-interacting protein, where the interac
 protein is chosen from trkA, protein phosphatase lalpa, 14-3-3ef
 alpha-tropomyosin, vimentin, p0071, Inl-1, IP-1, IP-2, IP-3, IP-4
 5. NIK1 (also referred to as Nek2) is a human homologue of the
 Aspergillus nidulans mitotic regulator, NIMA kinase. NIK1 is a
 serine/threonine-specific kinase and is thought to play a key rol
 cell-cycle events leading to the onset of mitosis. The complexes,
 derivatives and NIK1 or NIK1-IP protein and DNA sequences, etc. a
 useful for treating or preventing a disease or disorder involving
 aberrant levels of the complex or protein. Such disorders include
 hyperproliferative disorders, neurodegenerative disorders,
 cardiomyopathies, viral infections and metabolic disorders
 Sequence 1211 AA;
 Query Match 44.2%; Score 2818.5; DB 3; Length 1211;
 Best Local Similarity 49.2%; Pred. No. 3.1e-192;
 Matches 618; Conservative 192; Mismatches 337; Indels 109; C
 QY 15 MPVDPDPSASEKTSLSPLNTSGDSETEIT-SAILASVKEQELQFRLTREI
 DB 1 MPAPEQSLVEEGQFQTRQEA-ASTGCMPEPTATTILASVKEQELQFRLTREI
 QY 74 QIVASQLERCKLGSSTGSMSSSAEQFQWQSQD-----GQKDIEDELTTGLEI
 DB 60 QIVASQLERCKLGSSTGSMSSSAEQFQWQSQD-----GQKDIEDELTTGLEI
 QY 128 IRSIQESGIL-DPODYTGTERPSLL--SQSALQNSKPEGSFOYPASVHSNQTIA-
 DB 116 IRTEPEQGLYSPEQTSLSHESEGLSGNSRSTQMNYSQSDGYQEGAGSFHNSQNSK
 QY 180 -----LGETTPSQLPARGTQARATQSFQ-GTTSRAGHLAGPAPPPPP-----
 DB 176 QQHSFIGSTNNHV-----RNSRAGQTLVDPVANRARRVSSVPSRAQSPSYIS
 QY 225 PPREFPAPSLGSAFHLPP---DAPPAALYSSSTLPAPPRGSGSLAAPQGGSE
 DB 232 PSRSLRTSLGSGFGSPVTDPRPLNPSA---YSSITLPA-ARASPY-SQRPASE

3SAPEGATVAA PRGSPBK-QSPSR LAKSYSTSPINIVVSSAGLSPIRVTSPPTVQST 340
3SVTSRQT-SNPNGPFTQYQTTRV-----GSGFLTITDAQT-----RVASPSQGV-V 333
3SPIHOLSSTIGTYATLSPT-KRLVHAS EYQSKHSOELNATATLQRPGLSILAGSRASY 399
3SP--KRSGMTAVPOHLGSPSLQRTVHDMEGFGQQQYDIYERVMVPPRDSL-TGLRSY 390
3HGLGPELRALQSPHEHHIDPIYEDRVYQKPPMRSLSQSGDPLPAHTGTYRTSTAP 459
3HSOLQDRLRSVSPDLHITPIYEGRTYSPVYRSPNHGIVE-LOGSQATALYRT---- 445
3GVDSV-PIORTGSHGPQVAAAATFORASYAAGPASINADPYRQLOYCESVESPYSK 518
3GVSFGNLTQRTSSQR-----SLTYQRNNYALNTTATYAEPRYPYQYRVQ-ECNTNR 496
3ALPPEGTTLARSPSIDSIQKDPREFGWRDPPEVITOMLQHOQFSPVSQNAALVLOHLC 578
3AVPADDGTTSPSIDSIQKDPREFARWRDPPEVITMLHEHQFSPVQANAAALVLOHLC 556
3NKKIAEIRROGGIQLLVLDLDRHMTVEHRSACGALRNLVYKGANDDNKIALKNCGGI 638
3NKVKMEVCRLGGIKHLVDLLDHRVLEVOQKACGALRNLVFGKSTDENKIAMKNVGGI 616
3VLLRKTITDELRELTGVITWNLSSCDALKNPIIODALAVLTNAVII PHSGWENSPL 698
3LRLRKSIDA EVELRVTGVITWNLSSCDVAKMTIIRDALSTLTNTVIVPHSGWNSSF 676
3RKIQLHSSQVLNRNATGCLURNVSSAGEEARRRMWECGDTLLVYIQSALGSSEIDS 758
3HKIKFQTSVLVRNTTGCULRNTSAGEEARKQNRSCGIVDSLLVYIHTCVNTSDYDS 736
3ENCVCILRNLSYRLAAETSQOCHMGTDLDGLLCEANGKDAESSCGKKKKKKKK- 817
3ENCVCILRNLSYRLEVPQARLLGNELDDLKGESPSKQSEPS-CWGKKKKKKKR 795
3QDQWQGVGPLDCAEPPKGIQMLWHPSIVKPYLITLLSCSNPDITLGAAGALQNLAA 875
3JEDQWQGVGPIPLGSKPGVEMLWHPSVYKPYLITLLAESNPATLGEAGSLQNLUSA 855
3KWSVYIRAAVKEKGLPILVELLRITDNRVVCATARNMALDVRNKELIGKYAMR 935
3KFAAYIRGRPRKGLPILVELLRMDNRVSSGATARNMALDVRNKELIGKYAMR 915
3HRLPGGNNSNNTASKMSDDTVTAVCCTLHEVITKQNMENAKALRDAGGIEKLVGTSK 995
3NRLPGGNGPS-----VLSDETMAALCCALHEVTSKMENAKALADSGGIEKLVIITK 970
3DKHSPKVVKAASOVILNSMWQYRDLRSLYKKGQWGOYHFVASSSTIERDQRDPYSSR 1055
3DRSLKVVYKAAQVNLTIWQYRDLRSIYKKGQWGNHFITPVSTLERDFKSH----- 1026
3ISPV-RVSPNNRSASAPASPREMISLKERKTDVECTGSNATYHGAKGEHTSRKDM 1113
3LSTTNQMSPIIQSVGSTSSPALLGINDPRSEDRTOPPMQYNSQGDATHK--GL 1083
3NTGISTLYRNSYGAPAE-----IKHNQVSAQFVPEPSRKQYETYPQFQNSRNYDE 1169
3SSKSPYIYSSYSPAREQNRRLQHOQLYYS--QDDSNRKNFDAYRLVQLSPHSYED 1141
3EDQVHHRPPASEYTHMLGKSTGNVVDYFSAARPYSELNYTSTHYHPASPDGW 1225
3DRVHF-PASTDYQYGLKSTNYVDYFSTKRP-----SYRAEQYVPGSPDGW 1192

undard; protein; 1211 AA.

(first entry)

Human p0071 protein sequence SEQ ID NO:487.

Human; colon cancer; cancer; tissue profiling; forensic; mapping genetic analysis; diagnostic; antisense therapy.

Homo sapiens.

WC200229086-A2

11-APR-2002

[illegible]

1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

(PAGE / DATA CONT.)

buigess C, Astre JH, Carroll E, Cacho IS, Dwiwed F, Moll
Thiaglingam A, Lewis ME;

WPI: 2002-426115/45

N-PSDB; ABQ607B4.

New isolated nucleic acid that is differentially expressed in colon cancer tissues useful for determining the presence of colon cancer in a tissue type, and in antisense therapy.

Claim 5; Fig 3; 796pp; English.

ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially expressed in cancer tissues. ABQ78993 to ABQ79004 represent proteins encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) used in antisense therapy. An antibody immunoreactive with a protein encoded by (I) is useful for detecting cancer in a patient sample for detecting the presence or absence of a polynucleotide encoding a nucleic acid which hybridises to (I) in a cell. A probe/primer from (I) can be used for determining the presence of a nucleic acid which hybridises to (I), and for determining the phenotype of cells in culture of cells from a patient. (I) is useful for determining the presence of colon cancer in a cell or tissue type, for determining the presence or state of other type of cancer, in antisense therapy, to generate macroarrays on a solid surface, to identify a chromosome on which corresponding gene resides, and in tissue profiling, forensics, analysis, mapping and diagnostic applications. (I) can be used as antibodies, and to screen for peptide analogues and antagonists.

Sequence 1211 AA;

Query Match 44.2%; Score 2818.5; DB 5; Length 1211;
Best Local Similarity 49.2%; Pred. No. 3.1e-192;
Matches 618; Conservative 192; Mismatches 337; Indels 109;

15 MPVPDQPSSASEKTSLSPLNTSNGDSETETT-SAILASVKEQELQFERLTR

I MPAPQASLVEEGGPQTRQEA-ASTGPGMEPTATTILASVKEQEQLQFRLTR
74 QIVASOLERCKTSGTGTGSMSSMSAAEPQWQSOD-----GKKDIEDLETGGI

60 QIVASQLERCLGAESPISASTSTKSFPPWRSTDVPNTGVSKPRVSDAVQPNN:

```
128 IRSLQESGIL-DPQDYSTGERPSLL--SQSALQLNSKPEGSFQYPASYHSNQTLA
```

180 -----LGETTSQLPARGTQARATGQSFQ-GTTSRAGHLAGPEPAPPPPP-----

176 QQHSFIGSTNNHV-----RNSRAEGQTLVQPSVANRAMRRVSSVPRAQSPSYV

225 PPREPAPSLGSAFHP---DAPPAANAALYYSSSTLPAPPRGGSPLAAPQGG
232 DSPGCSIPTSICSGCGSGDSVTNDDDFNDCA---VCSSTTIDAAADACDV-SQBBAC

282 PCCSADPCATVADPCCSDPK-CCSDP1AKVGVCTCCPTNTVWCCACTCSDPTWTCDBI


```

VTSSQT-SNPNGPPQYQTTARV-----GSPLLTLTAQT-----RVASPSQGV-V 333
PIHQLSSITIGYATILSPT-KRLVHASQYKSHSQELVATATLQRPGLSAGSRASY 399
P--KRSGMTAVPQHLGFSQRTVHMEQFGQQQYDIYERWVPPRPDSL-TGLRSSY 390
GHLGPELRALQSPHEHIDPIYEDRVYQKPPMRSLSSQSGDPLPPAHTTGYTSTAP 459
SQLQDLRSVSPDLHTPIYEGRTYSPVYRSPNHTGIVE-LQSQTALYIT---- 445
VDSV-PLQRTGSHQHPQNAATFORASYAAGPASNYADPYRLQYCPSVESPYSK 518
VSGIGNLQRTSSQR-----STLTQYNNYALNTTATYAEPYRPIQYRVQ-ECNYNR 496
LPPEGTLARSPSIDSIQKDPREFGWRDPELPEVQMLQHPSPVSQSNAAVLOHLC 578
VPADDGTRSPSIDSIQKDPREFAWRDPELPEVHMLEHOFPSVOANAAVLOHLC 556
KIKAEIRROGGIQLLVLLDHRMTVRHSACALRNLYVKANDDNKIALKNCGGI 638
KVKMEVCLGGIKHLVDLLDHRVLEQVKNACGALRNLYVFGKSTDENKIAMKNVGGI 616
RLLRKTTDLREIRLVTVGLWNLSSCDALKMPIIODALAVLTNAVITPHSGWENSGPL 698
RLLRKSIDAEVRELVTGLWNLSSCDVAKMTIIRDALSTLTNTVIVPHSGWNNSSF 676
KIQHSSQVLTNATGCLRNVSAGAEARMRBCDGLTDALLYVTQSLGSEIDS 758
IKIKFTSLVLRNTTCLRNLYSAGEEARQMRSCGLVDLSLLYVHTCVNTSDYDS 736
NCVCTLRNLSYRLAAETSGQHMGTDELDGLLCEANGKDAESSQGWKKKKKKK- 817
NCVCTLRNLSYRLELEVQARLLGLNELDGLLGKSPSKDSEPS-CWKKKKKKKKR 795
DQWDGVGLPDCAEPPKGIOMLWHPISYKPYLTLLSECSNPDTLEGAGALONLAA 875
DQWDGVGPIPLGLSKSPKGVEMLWHPISYKPYLTLLAESNPATLBSAGSLQNLAA 855
WSVYIRAAVRKEKGIPIIVLELLRIDNDRVCAVATARNMALDVNRELIGKYAMR 935
FAAYIRGGRPKRGKGIPIIVLELLRMONDRVSSGATARNMALDVNRELIGKYAMR 915
IRLPGNNNSNTASKAMSDDTTAVCCTLHEVITKNMENAKALRDAGGIEKLVGISK 995
RLPGNGGFS-----VLSDETMAAICCALHEVTSKNMENAKALADSGGIEKLVNITK 970
KHPKPVVKAASQVLNSMWQYRDLRLSLYKDGWSQYHFVASSSTIBDRQRPSYSSR 1055
RSSLKVVKAAQAQVLTNLWQYRDLRLSLYKDGWNQNHFTTPVSTLERDRFKSH- 1026
SPV--RVSPNNRSAPAPREMI SILKERTDYECTGSGNATVHGAKGHTSRKDM 1113
STTNQOMPIQSVGCTSSSPALLGIRPRSEYDQTPPMQYNSQGDATHK--GL 1083
TGISTLYRNSYGAPED----IKHNQVSAQVPQPSKDYETYPQFNSTNRNDE 1169
SKPSPIYISSYSSPAREQNRRLQHQQLVYS--QDDSNRKNFADYRLYLQSPHSYED 1141
DOVHRHPASBYTHMLGLKSTGNYDYFSAAPPYSELNLYETSHYPASPDSSW 1225
DRVHF-PASTDYSTQYGLKSTTNYDYFSTRKP-----SYRAEPQYGPSPDSW 1192

```

idard; protein; 1110 AA.

(first entry)

llular matrix and cell adhesion molecule-12 (XMAD-12).

XX Human; extracellular matrix and cell adhesion molecule; XMAD;
 KW gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
 KW Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dy;
 KW sickle cell anaemia; thalassaemia; autoimmune disorder; adenocar;
 KW inflammatory disorder; acquired immune deficiency syndrome; AIDS
 KW Addison's disease; allergy; anaemia; asthma; atherosclerosis; me;
 KW Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma;
 KW glomerulonephritis; multiple sclerosis; Grave's disease; osteoar;
 KW osteoporosis; psoriasis; rheumatoid arthritis; ulcerative coliti;
 KW infection; cell proliferative disorder; actinic keratosis; myelo;
 KW arteriosclerosis; nontropic; anticonvulsant; antithyroid; nephroi;
 KW neuroprotective; dermatological.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FH 511..1025 /note= "Mouse p120 protein"
 FT Region
 FT 532..565 /note= "Armadillo/beta-catenin-like repeats"
 FT 577..598 /note= "Armadillo/beta-catenin-like repeats"
 FT 609..622 /note= "Armadillo/beta-catenin-like repeats"
 FT 971..973 /note= "Rgd cell interaction motif"
 FT WO200142285-A2.
 XX 14-JUN-2001.
 XX 05-DEC-2000; 2000WO-US032990.
 XX 10-DEC-1999; 99US-0172852P.
 XX 16-DEC-1999; 99US-0172354P.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;
 PI Baughn MR, Lu DAM, Shah P, Au-Young J;
 PI WPI; 2001-381632/40.
 DR N-PSDB; AAD08056.
 XX New human extracellular matrix and cell adhesion molecules and
 PT polynucleotide sequences encoding them, useful for diagnosis, pr
 PT treatment of genetic, autoimmune and cell proliferative disorder.
 XX Claim 1; Page 104-107; 135pp; English.
 XX The present sequence is a human extracellular matrix and cell ad
 CC molecule (XMAD). The XMAD is used for screening a compound for
 CC effectiveness as an agonist or antagonist of XMAD. The identifie
 CC or antagonist are used for treating a disease or condition assoc
 CC with decreased or increased expression of functional XMAD. The
 CC polynucleotides encoding XMAD are useful in somatic or germ-line
 CC therapy to correct a genetic deficiency, to express a conditiona
 CC lethal gene product and to express a protein which affords prote
 CC against intracellular parasites and also for diagnosis of disord
 CC associated with expression of XMAD. They are also used for gener
 CC hybridisation probes useful in mapping the naturally occurring g
 CC sequences and to create knock in humanised animals (pigs) or tra
 CC animals (mice or rats) to model human diseases. Oligonucleotide
 CC fragments derived from the polynucleotide sequences may be used
 CC elements on a microarray. Antibodies which specifically bind XMAD
 CC used for the diagnosis of disorders associated with the expressio
 CC XMAD, or in assays to monitor patients being treated with XMAD. I
 CC diagnosed, prevented or treated include genetic disorders such a
 CC adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher'
 CC disease, myotonic dyetrophy, sickle cell anaemia, thalassaemia,
 CC autoimmune/inflammatory disorders such as acquired immune defici
 CC syndrome (AIDS), Addison's disease, allergies, anaemia, asthma,

Qy	877	SWKWSVYIRAAVKRKGKGLPILVELLRINDRVVCATAFALRNMAALDVNRKELIG	
Db	856	NWKFAYTRAAVKRKGKGLPILVELLRMDNDRVSVATAFALRNMAALDVNRKELIG	
Qy	937	LVHRLPGGNSNNTASKAMSDDTTAVCTTLEHVTTKMENAKALRDAGGIEKL	
Db	916	LVNRLPGGNGPS-----VLSDETMAAICAHLEVTSKMENAKALADSGGIEKL	
Qy	997	KGDGSKPKVVKAAASQVLNSMMQYRDLRLSYKKDGMSQYHFVASSSTIERDRQRP	
Db	971	RGDRSSLKVVKAAQAQLNTLMQYRDLRLSYKKDGWNQNHFTTPVSTTLERDRFKSI	
Qy	1057	PSISPV--RVSPNNBSASAPSPREWSILKERRKTDYECTGSGNATYHGAKGBHTSI	
Db	1026	PSLSTTNQOMSPFIQSVGSTSSPALLGIRPRSEYDRTPQPMQYNSQGDATH	
RESULT	12		
ABP	41836		
ID	ABP41836	standard; protein; 785 AA.	
XX	ABP41836;		
AC	ABP41836;		
DT	22-AUG-2002	(first entry)	
XX			
DE		Human ovarian antigen HOCPT34, SEQ ID NO:2968.	
XX			
KW		Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour; ovarian cancer; breast cancer; tumour; reproductive system disorder; infertility; pregnancy disorder; anovulation; polycystic ovary; PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infectious inflammatory condition; immune disorder; blood disorder; cardiovascular disorder; respiratory disorder; neurological disorder; gastrointestinal disorder; urinary system disorder; drug screen; gene therapy; chromosome mapping; forensic analysis; antibody preparation; cytostatic; immunomodulatory; neuroprotectant; inflammatory; gynaecological; reproductive; chromosome 2q23.	
XX			
OS		homo sapiens.	
XX			
PN		WQ200200677-A1.	
XX			
PD		03-JAN-2002.	
XX			
PF		07-JUN-2001; 2001WO-US018569.	
XX			
PR		07-JUN-2000; 2000US-0209467P.	
XX			
PA		(HUMA-) HUMAN GENOME SCI INC.	
XX			
PI		Birse CE, Rosen CA;	
XX			
DR		WPI; 2002-147878/19.	
DR		N-PSDB; ABQ54913.	
XX			
PT		Isolated nucleic acid molecules encoding novel ovarian polypeptide useful in the prevention, treatment and diagnosis of cancer (e.g. cancer), immune disorders, cardiovascular disorders and neurological diseases.	
XX			
PS		Claim 11; SEQ ID NO 2968; 2922pp; English.	
XX			
CC		The invention relates to 2175 novel human ovarian antigens (ABP: ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and encompasses polypeptides 90% identical and polynucleotides 95% identical to the sequences of the invention. The invention additionally relates to recombinant vectors and host cells comprising human ovarian antigen polynucleotides, antibodies against human ovarian antigens, and of ovarian antigen polynucleotides and polypeptides in diagnosis of treating, prognosing or preventing various ovary and/or breast disorders. Such conditions include ovarian cancer and breast cancer metastatic tumours of ovarian or breast origin, reproductive system	

```

QY      1046 DRQPYSSRTPSPV--RVSPNNRSASAPAPREMI SLKERKDYECTGSGNAT.
          ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      705 DRFKSH-----PSLSTTNQOMSPIOSVGSTSSPALLGIIDRPSEYDRTQPPMQ
          ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      1104 GEHTSR 1109
          ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB      760 GDATHK 765
          ||| : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 13
ABG04995
ID ABG04995 standard; protein; 696 AA.
XX
XX
AC ABG04995;
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #4986.
DE
```

XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
XW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
XW	
XX	Homo sapiens.
XW	
XX	
XW	

PN WO200175067-A2.
XX

1	1
2	6
3	5
4	6
5	6
6	5
7	2
8	6
9	6
10	6
11	6
12	6
13	6
14	6
15	6
16	6
17	6
18	6
19	6
20	6
21	6
22	6
23	6
24	6
25	6
26	6
27	6
28	6
29	6
30	6
31	6
32	6
33	6
34	6
35	6
36	6
37	6
38	6
39	6
40	6
41	6
42	6
43	6
44	6
45	6
46	6
47	6
48	6
49	6
50	6
51	6
52	6
53	6
54	6
55	6
56	6
57	6
58	6
59	6
60	6
61	6
62	6
63	6
64	6
65	6
66	6
67	6
68	6
69	6
70	6
71	6
72	6
73	6
74	6
75	6
76	6
77	6
78	6
79	6
80	6
81	6
82	6
83	6
84	6
85	6
86	6
87	6
88	6
89	6
90	6
91	6
92	6
93	6
94	6
95	6
96	6
97	6
98	6
99	6
100	6

XX
PR 31-MAR-2000: 2000US-00540217.

FR 23-AUG-2000; 2000US-00649167.
XX
XX

[illegible]

XX
DR WPI: 2001-639362/73.

DR N-PSDE; AAS69182. XX

PT diagnostics, forensics, gene mapping, identification of mutations

PT biodiversity.
XX

XX
XX

CC sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping.

CC and in recombinant production of (11). The polynucleotides are a

CC in diagnostics as expressed sequence tags for identifying expressed

CC activity of (II) or to treat disease states involving (II). (II)

CC polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical

CC involving aberrant protein expression or biological activity. The
CC of sites expressing (I) and (II) are useful for treating a

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodi-

CC and to produce other types of data and products dependent on DNA
CC amino acid sequences. ABG00010-ABG30377 represent novel human dia

CC amino acid sequences of the invention. Note: the sequence data in
CC patent did not appear in the printed specification, but was obtained

ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 696 AA;

אשר ימלאו
33.9% ; 3001 אשר
במ , 1, אשר
אשר , אשר

larity 78.9%; Pred. No. 9.2e-155;
Conservative 13; Mismatches 19; Indels 90; Gaps 3;

EQFQWSDGQXDIEDELTTGLELVDSICIRSQESGIILDPQDYSTGER-----147
EQFQWSDGQXDIEDELTTGLELVDSICIRSQESGIILDPQDYSTGERVIHFGKTLG120
-----147

EDYSIMEXQMVKVETVTVPQSVLKANETMASALCGQEHDPLLLTVQMSTGNHYI180
-----PSLLSQSALQLNSKPESFYQPASYHSNQTIALGETTPSQLPARGTOAR196
LSLFPPWKVPSLLSQSALQLNSKPESFYQPASYHSNQTIALGETTPSQLPARGTOAR240
QSFSQGTTSRAGHLGAPPEAPPPPPPFRPPFPAPSIGSAFHLPDAPPAAAAAAYYSS256
JAPPRGGSPLAAPQGSGPTKLORGSAPEGATYAAPRGSSPKQSPSRSLAKSYSTSSP316
JAPPRGGSPLAAPQGSGPTKLORGSAPEGATYAAPRGSSPKQSPSRSLAKSYSTSSP360
VVSSAGLSPIRVTSPTTVOSTISSSPIHQLSSTIGTYATLSPTKKLVHASEQYSKHS376
VVSSAGLSPIRVTSPTTVOSTISSSPIHQLSSTIGTYATLSPTKKLVHASEQYSKHS420
XYATLILORPGSLAAGSRASYSQHGHGLPELRALQSPHHIDPIYEDRYVQKPPMRS436
XYATLILORPGSLAAGSRASYSQHGHGLPELRALQSPHHIDPIYEDRYVQKPPMRS480
XSQGDPLPPAHTGTYTISTAPSSPGVDSPVLQRTGSQHGPQNAAAATFORASYAAGPA496
XSQGDPLPPAHTGTYTISTAPSSPGVDSPVLQRTGSQHGPQNAAAATFORASYAAGPA540
ADPYRLQLQCYCPVESPYSKSGPALPPEGTLARSPSIDSIQKDPRFF----GWRDDEL552
ADPYRLQLQCYCPVESPYSKSGPALPPEGTLARSPSIDSIQKDPSKYPILCDGNGNPF600

/IQMLQHQPFSVGSNAAAYLQHLCFGDNKIKAEIRR590
/YR-----PSSRRERASSLLKSMCYQXNRLYPKERR632

undard; protein; 1009 AA.

(first entry)

diagnostic protein #20811.

nosome mapping; gene mapping; gene therapy; forensic;
nent; medical imaging; diagnostic; genetic disorder.

S.

-A2.

; 2001WO-US008631.
; 2000US-00540217.
; 2000US-00649167.
3Q INC.
Liu C, Tang YT;
9362/73.

ILLEAYEFATSVLAGSSVSTLPHLLAPKLIHAWSLAERGRKSEAQQYCDIAAAL 1311
XSKG-----DKHSPKVVKAASQVLNMMQYRDLRLSLYKKG--WSQVH-FV 1037
KPSGYHNOHLFFGVDELARLRETTSDG-GSSWISRP--SWEKVSGSMWAKFNSFV 1368
---ASSSTIERDRQRPYSS-----SRTPSIS-----PVRVS 1065
SDAASTGSKVAEBIGPFARYSGTPTISRSFVSVDIYGSYPVAAAQPLPATGPSRYQ 1428
SASAPASPREMISLKERKTDYECTGNSATYHGAKGEHTSRKDAMTAQNTGI----- 1120
ZAPS-ASPEQL-----RGRSSMDSQRSASFYPLGQRGSOEESTPYDTNMFHGM 1482
-----STLYRNSY-----GAPAEIDK-HNOVSAQVPQEPSPKDYET 1156
PVAGYQSTPPQSSYMLAPVAEDSASGAQOESFSAHSQVS-----DNAPSHRS-ST 1537
PQON--STRNYDESFEDQVHHRPPAS 1183
FGHPEDTQAVSTTSQPDQGGYMEPTS 1568

April 22, 2004, 12:21:03


```

|||||
GTYTSTAPSPGVDSVPLQRTGSGRGNAAAATFQASAAAGPASYADPYQLQ 366
SVESPYSGPALPEGTTLARSPSIDSKQDPREFGWRDPELPEVIQMLQHFPSVQ 567
SVESPYSGPALPEGTTLARSPSIDSKQDPREFGWRDPELPEVIQMLQHFPSVQ 426
AAYLQHLGFCGDNKIKAETRRGGTQLLVLDLDRHMTVEVRSACGALRLVYKANDD 627
AAYLQHLGFCGDNKIKAETRRGGTQLLVLDLDRHMTVEVRSACGALRLVYKANDD 486
ALKNCGGIPALVRLLRKTTDLLEIRELVGLVNLSSCDALKMPILQDALAVLTNAVI 687
ALKNCGGIPALVRLLRKTTDLLEIRELVGLVNLSSCDALKMPILQDALAVLTNAVI 546
ISGWENSPLODDRKIQLHSSQVLRNATGCLRVSSAGEEARRMRRECDGLTDALLYVI 747
ISGWENSPLODDRKIQLHSSQVLRNATGCLRVSSAGEEARRMRRECDGLTDALLYVI 606
LGSSEIDSKTVENCVCILRLNLSYLAETSQGHMTDDELGLLCEANGKDAESSG 807
LGSSEIDSKTVENCVCILRLNLSYLAETSQGHMTDDELGLLCEANGKDAESSG 666
KKKKKKKSQDQDVGVLPCDAEPPKGIQMLWHPISVVKPYLTLLSECSNPDTLGGAA 867
KKKKKKKSQDQDVGVLPCDAEPPKGIQMLWHPISVVKPYLTLLSECSNPDTLGGAA 726
QNLAAAGSKWQSVYIRAAVRKEKGLPILVELLRIDNDRVVCATARNMALDVRENKE 927
QNLAAAGSKWQSVYIRAAVRKEKGLPILVELLRIDNDRVVCATARNMALDVRENKE 786
KYAMRDLVHRLPGGNNNNNTASKAMDDTVTAVCCTLHEVITKMNENAKALRDAGGI 987
KYAMRDLVHRLPGGNNNNNTASKAMDDTVTAVCCTLHEVITKMNENAKALRDAGGI 846
VGISKSGDKHSPKVVKAAASQVLSNMWQYRDLRLSLYKKGWSQVHFVASSSTIERDR 1047
VGISKSGDKHSPKVVKAAASQVLSNMWQYRDLRLSLYKKGWSQVHFVASSSTIERDR 906
YSSRTFSPISPRVSPNNRSASAPSPREMISLKERKTDYECTGNSNATYHGAKGEHT 1107
YSSRTFSPISPRVSPNNRSASAPSPREMISLKERKTDYECTGNSNATYHGAKGEHT 966
DAMTAQNTGISTLRNSYGAPAEIDIKHNOVSAQVPQEPSPKDYETYPFQFONSTRNY 1167
DAMTAQNTGISTLRNSYGAPAEIDIKHNOVSAQVPQEPSPKDYETYPFQFONSTRNY 1026
FFEDQVHRRPPASBYTHHLGLKSTGNVYDFYSAARPYSELNYETSHYPASPDSMW 1225
FFEDQVHRRPPASBYTHHLGLKSTGNVYDFYSAARPYSELNYETSHYPASPDSMW 1084

```

lication US/08982785A

929

ATION:
Kosik, Kenneth S.
Zhou, Jianhua

ENTION: ALARM RELATED PEPTIDES AND
ENTION: NUCLEIC ACIDS AND DIAGNOSIS USING THEM

QUENCES: 13

CE ADDRESS:

Fish & Richardson P.C.
25 Franklin Street
ton

USA

0-2804

DABLE FORM:

E: Diskette

IBM Compatible

SYSTEM: Windows 95

SOFTWARE: FastSeq for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/982,785A

FILING DATE: 02-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/031,556

FILING DATE: 02-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Freeman, John W.

REGISTRATION NUMBER: 29,066

REFERENCE/DOCKET NUMBER: 05311/018001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 756 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

US-08-982-785A-2

Query Match 61.4%; Score 3916; DB 3; Length 756;
Best Local Similarity 94.9%; Pred. No. 4.2e-277;
Matches 752; Conservative 0; Mismatches 4; Indels 36;

```

QY 434 MRSLSSQSGDPLPAHTGTVRTSTAPSSPGVDSVPLQRTGSGRGNAAAATFQF
Db 1 MRSLSSQSGVPLPAHTGTVRTSTAPSSPGVDSVPLQRTGSGRGNAAAATFQF
QY 494 GPASNADPYRQYQCPSPSPSKSGPALPPEGTTLARSPSIDSIQKDPREFGWF
Db 61 GPASNADPYRQYQCPSPSPSKSGPALPPEGTTLARSPSIDSIQKDPREFGWF
QY 554 EVTOMLQHPSPVSNAAVLAHLCEGDNKIKAETRRGGTQLLVLDLDRHMTVE
Db 121 EVTOMLQHPSPVSNAAVLAHLCEGDNKIKAETRRGGTQLLVLDLDRHMTVE
QY 614 GALRLNVYKANDDNKIATLKNCGGIPALVRLLRKTTDLLEIRELVTVGLVNLSSCI
Db 181 GALRLNVYKANDDNKIATLKNCGGIPALVRLLRKTTDLLEIRELVTVGLVNLSSCI
QY 674 IIQDALAVLTNAVIIIPHSGWENSPLODDRKIQLHSSQVLRNATGCLRVSSAGEE
Db 241 IIQDALAVLTNAVIIIPHSGWENSPLODDRKIQLHSSQVLRNATGCLRVSSAGEE
QY 734 RECDGLTDALLYVIQSLAGSSEIDSKTVENCVCILRLNLSYLAETSQGHMGTI
Db 301 RECDGLTDALLYVIQSLAGSSEIDSKTVENCVCILRLNLSYLAETSQGHMGTI
QY 794 LCGEANGKDAESSGCGWKKKKKKKSDQDQDVGVLPCDAEPPKGIQMLWHPISVVK
Db 361 LCGEANGKDAESSGCGWKKKKKKKSDQDQDVGVLPCDAEPPKGIQMLWHPISVVK
QY 854 LSECSNPDTLEGAAGALQNLAAAGSKWQSVYIRAAVRKEKGLPILVELLRIDNDRV
Db 421 LSECSNPDTLEGAAGALQNLAAAGSKWQSVYIRAAVRKEKGLPILVELLRIDNDRV
QY 914 TALRNALDVNRKELIGKYAMRDLVHRLPGGNNNNNTASKAMSDDDTVTAVCCTLH
Db 481 TALRNALDVNRKELIGKYAMRDLVHRLPGGNNNNNTASKAMSDDDTVTAVCCTLH
QY 974 NMENAKALRDAGGIEKLVGISKSGDKHSPKVVKAAASQVLSNMWQYRDLRLSLYKK
Db 541 NMENAKALRDAGGIEKLVGISKSGDKHSPKVVKAAASQVLSNMWQYRDLRLSLYKK
QY 1034 YHPVASSSTIERDRQRPYSSSRTPSISPRVSPNNRSASAPASPREMISLKERKT
Db 601 YHPVASSSTIERDRQRPYSSSRTPSISPRVSPNNRSASAPASPREMISLKERKT
QY 1094 GSNATYHGAKGEHTSRKDAWTAQNTGISTLRNSYGAPAEIDIKHNOVSAQVPQFQ

```


|||||
TYHGAKGHTSRKDAQNTGISTLYR----- 692
QPFCNSTRNDESEFFEDQVHRPPASEYTMHLGKSTGNVYDFYSAARPYSELNVE 1213
-----NSTNDESEFFEDQVHRPPASEYTMHLGKSTGNVYDFYSAARPYSELNVE 744
PASPSDW 1225
PASPSDW 756
ication US/08982785A
29
TION:
ou, Kenneth S.
hou, Jianhua
TION: ALARM RELATED PEPTIDES AND
UTION: NUCLEIC ACIDS AND DIAGNOSIS USING THEM
UNCES: 13
E ADDRESS:
Fish & Richardson P.C.
5 Franklin Street
on
SA
-2804
ABLE FORM:
: Diskette
IBM Compatible
YSTEM: Windows 95
FastSEQ for Windows Version 2.0b
CATION DATA:
NUMBER: US/08/982,785A
: 02-DEC-1997
TION DATA:
NUMBER: US 60/031,556
: 02-DEC-1996
I INFORMATION:
nan, John W.
N NUMBER: 29,066
CKET NUMBER: 05311/018001
TION INFORMATION:
617/542-5070
17/542-8906
154
SEQ ID NO: 8:
ACTERISTICS:
6 amino acids
c acid
linear
: protein
55.8%; Score 3558; DB 3; Length 686;
arity 99.4%; Pred. No. 4.7e-251;
onservative 0; Mismatches 4; Indels 0; Gaps 0;
3DLPAPHTTYRTSTAPSPGVDSVPLQRTGSHQHPQNAATAATFORASYAAGPAS 497
GVPLPPAHTGTYRTSTAPSPGVDSVPLQRTGSHQHPQNAATAATFORASYAAGPAS 60
PYRLOQCPSPVSPYSKSGALPEGLTARSPSIDSIQKDPREFGWRDPPEVIQ 557
PYRLOQCPSPVSPYSKSGALPEGLTARSPSIDSIQKDPREFGWRDPPEVIQ 120
QFSPVQNAAYLQHLFCGDNKIKAEIRROGGIQLLVLDLDRHMTVEHRSACGALR 617
QFSPVQNAAYLQHLFCGDNKIKAEIRROGGIQLLVLDLDRHMTVEHRSACGALR 180
3KANDDNKIALKNCGGIPALVRLRKTDTLEIRELVTVGLVNLSSCDALMPLIQD 677

Db 181 NLVYKANDDNKIALKNCGGIPALVRLRKTDTLEIRELVTVGLVNLSSCDALMPLIQD 677
QY 678 ALAVLTNAVITPHSGWENSPLODDRKIQHSSQVLRNATGCLRNVSSAGEARRR
Db 241 ALAVLTNAVITPHSGWENSPLODDRKIQHSSQVLRNATGCLRNVSSAGEARRR
QY 738 GLTDALLYVVIQSALGSSSEIDSKTVENCVCILRNLSYRLAAETSOQHMGTDDELDCI
Db 301 GLTDALLYVVIQSALGSSSEIDSKTVENCVCILRNLSYRLAAETSOQHMGTDDELDCI
QY 798 ANGKDAESSGCGWKKKKKKKSDQDWDGVPPLDCAEPPKGIOMLWHPHSIVKPYLTI
Db 361 ANGKDAESSGCGWKKKKKKKSDQDWDGVPPLDCAEPPKGIOMLWHPHSIVKPYLTI
QY 858 SNPTDLGAGALONLAAGSMKWSVIYIRAAVRKEKGLPILVELLRINDRNVCAV
Db 421 SNPTDLGAGALONLAAGSMKWSVIYIRAAVRKEKGLPILVELLRINDRNVCAV
QY 918 NMALDVRNKELIGKYAMRDLVHRLPGGNNNSNTASKAMSDDTVTAVCCTLHEVIT
Db 481 NMALDVRNKELIGKYAMRDLVHRLPGGNNNSNTASKAMSDDTVTAVCCTLHEVIT
QY 978 AKALRDAGGIEKLVGISKSGDKHSPKVYVAAASQVLSNMMQYRDLRSLYKKDGWSC
Db 541 AKALRDAGGIEKLVGISKSGDKHSPKVYVAAASQVLSNMMQYRDLRSLYKKDGWSC
QY 1038 ASSSTIERDRQRPYSSSRTPTSPVSPVSPNRSASAPSPREMISLKERKTDYECT
Db 601 ASSSTIERDRQRPYSSSRTPTSPVSPVSPNRSASAPSPREMISLKERKTDYECT
QY 1098 TYHGAKGHTSRKDAQNTGISTL 1123
Db 661 TYHGAKGHTSRKDAQNTGISTL 686
RESULT 4
US-08-982-785A-10
; Sequence 10, Application US/08982785A
; Patent No. 6258929
; GENERAL INFORMATION:
; APPLICANT: Kosik, Kenneth S.
; APPLICANT: Zhao, Jianhua
; TITLE OF INVENTION: ALARM RELATED PEPTIDES AND
; TITLE OF INVENTION: NUCLEIC ACIDS AND DIAGNOSIS USING THEM
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982,785A
; FILING DATE: 02-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,556
; FILING DATE: 02-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 05311/018001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 10:

CHARACTERISTICS:
20 amino acids
no acid
linear
E: protein

47.5%; Score 3030; DB 3; Length 620;
Identity 95.6%; Pred. No. 1.4e-212;
Conservative 1; Mismatches 26; Indels 0; Gaps 0;

SQSGDPLPPAHTGYRTSTAPSPGVDSVPLORTGSHGQPNAAAATFORASYAAG 494
|||||
SQSGVPLPPAHTGYRTSTAPSPGVDSVPLQRTCSQHICQNAAAAATFORASYAAG 60
|||||
NYADPYRQLOQCPSPVESPYSGSPALPEGLTARSPSIDSIQKDPREFGRWDPELPE 554
|||||
NYADPYRQLOQCPSPVESPYSGSPALPEGLTARSPSIDSIQKDPREFGRWDPELPE 120
|||||
MLQHPFVSQNAAYIQLHCFGDNKIKAEIRROGGIQLLVLDLDRHMTVEHRSACG 614
|||||
MLQHPFVSQNAAYIQLHCFGDNKIKAEIRROGGIQLLVLDLDRHMTVEHRSACG 180
|||||
NLVYKANDNDKIALKNCGGIPALVRLLRKTTDLIEBELVTGVLWNLSDDLKMPI 674
|||||
NLVYKANDNDKIALKNCGGIPALVRLLRKTTDLIEBELVTGVLWNLSDDLKMPI 240
|||||
ALAVLTNAVVIIPHSGWENSPLODDRKIQHSSQVLRNATGLRNVSSAGBEARRMR 734
|||||
ALAVLTNAVVIIPHSGWENSPLODDRKIQHSSQVLRNATGLRNVSSAGBEARRMR 300
|||||
GLTDALLYVQISALGSSEIDSKTVENCVCILRNLSYRLAETSQGHMGTDGLDGL 794
|||||
GLTDALLYVQISALGSSEIDSKTVENCVCILRNLSYRLAETSQGHMGTDGLDGL 360
|||||
ANGKDAESSGCGWKKKKKSDODWGVPLDCAEPPKGFOMLWHPISVVKPYLTLL 854
|||||
ANCFDAESSCCWKKKKKSONQWGVPLDCAEPPKGFOMLWHPISVVKPYLTLL 420
|||||
SNPDTLEAGACALQNALAGSKWKSIVYIRAAVRKEKGLPILVELLRINDRNVCAVAT 914
|||||
SNPDTLEAGACALQNALAGSKWKSIVYIRAAVRKEKGLPILVELLRINDRNVCAVAT 480
|||||
NMALDVRNKELIGKYAMRDVLRPLPGNNNSNTASKAMSDDTVTAVCTLHEVITKN 974
|||||
NMALDVRNKELIGKYAMRDVLRPLPGNNNSNTASKAMSDDTVTAVCTLHEVITKN 540
|||||
AKALRDAGGIEKLVGISKSGDKHSPKVVKAAASQVLSNMWQYRDLRLSLYKKGWSQY 1034
|||||
AKALRDAGGIEKLVGISKSGDKHSPKVVKAAASQVLSNMWQYRDLRLSLYKKGWSQY 600
|||||
ASSTIERDRQRPY 1051
|||||
ASSTIERDRQRPY 617
|||||

Publication US/09167206A

193

193

193

193

193

193

193

193

193

193

193

193

193

193

193

193

193

193

193

193

193

193

ORGANISM: Homo sapiens
US-09-167-206-14

Query Match 44.2%; Score 2818.5; DB 4; Length 1211;
Best Local Similarity 49.2%; Pred. No. 1e-196;
Matches 618; Conservative 192; Mismatches 337; Indels 109;

QY 15 MPVPDQPSASASEKTSLSPLGNTSNGDGSSETTT-SAILASVKEQELQFERLTRE
|||
DB 1 MPAPQASLVEEQPQTRQEA-ASTGPGMEPETATITILASVKEQELQFERLTRE
|||
QY 74 QIVASOLERCKLGSETGSMSSMSABEQFOWOSQD-----GQKDIEDELTGLF
|||
DB 60 QIVASOLERCKLGSETGSMSSMSABEQFOWOSQD-----GQKDIEDELTGLF
|||
QY 128 IRLSLQSSGIL-DFQDYSTGTGERPSELL--SOSALQNSKPEGSFOYPASYHSNQTLA
|||
DB 116 IRTEPQGLIYSPQYSLHSESGSLGNSRSSTOMNSYSDSGYQEAQSFHNSQNVV
|||
QY 180 -----LGETTPSOLPARGTQARATGQSFQ-GTTSRAGHLAGDEPAPPPPP-----
|||
DB 176 QHSHFGISTNNHV-----RNSRAEGQTLVQPSVANRMRVSSVPSRAQSPSYVI
|||
QY 225 PPREPAPSLGSAFHLP---DAPPAALAAALYYSSSTLPAPPGRGSPPLAAPQCGS
|||
DB 232 PSRGSRLTSLGSGFGSPVTDPRPLNPSA---YSSTLTPA-ARAAAPY-SQRPAS
|||
QY 282 RGGSAPEGATYAAPRGSSPK-QSPSLAKSYSTSSPINIVVSSAGLSPIRVTSPP
|||
DB 287 RIGSVTSRQT-SNPNGPTQYQTTARV-----GSPLLTDAQT-----KVAFPS
|||
QY 341 ISSSPIHQLSSTIGTYATLSPT-KRLVHASEQYKSHSQELYATATLQRPQSLAAG
|||
DB 334 GSSSP-KRSGMTAVPQHLGSLQRTVHDMEQPGQQQYDIYERWVPPRDSL-TG
|||
QY 400 SSQGHGLGELRALQSPHIDPIYEDRVYQKPPMRSLSSQSQDQDPIPPAHTGYR
|||
DB 391 ASQHSQGLQDLSAVSPDLHTPIYEGRTYYSVYRSPNHGTVE-LQGSOTALYR
|||
QY 460 SSPGVDSV-PLQRTSGHGPQNAATAATFORASYAAGPASNADPYRQLOQCPSPVE
|||
DB 446 ---GVSGIGNLQRTSSQR-----STLTQNNALNTTATYAEPIQIRVQ-E
|||
QY 519 SGPALPEGTTLARSPSIDSIQKDPREFGRWDPELPEVIQMLQHPFVSQNAAY
|||
DB 497 LQHAVPADDTTSPSIDSIQKDPREFGRWDPELPEVIQMLQHPFVSQNAAY
|||
QY 579 FGDNKKABIRQGGTQLLVLDLDRHMTVEHRSACGALRNLYVYGKANDNDKIALK
|||
DB 557 FGDNKKABIRQGGTQLLVLDLDRHMTVEHRSACGALRNLYVYGKANDNDKIALK
|||
QY 639 PALVRLLRKTTDLIEBELVTGVLWNLSDDLKMPIIODALAVLTNAVVIIPHSGW
|||
DB 617 PALVRLLRKTTDLIEBELVTGVLWNLSDDLKMPIIODALAVLTNAVVIIPHSGW
|||
QY 699 QDRKIQHLSQVLRNATGLRNVSSAGBEARRMRRECGLTDALLYVQISALGS
|||
DB 677 DDDHKIKFQTSILVLRNTTGLRNLTSAGEARKQMRSCGELVDSLLYVHTCVNT
|||
QY 759 KTVENCVCILRNLSYRLAETSQGHMGTDGLDGLCGFANGKDAESSCCWKKK
|||
DB 737 KTVENCVCILRNLSYRLAETSQGHMGTDGLDGLCGFANGKDAESSCCWKKK
|||
QY 818 --SQDQWDGVPGLPDCAPPKGIQMLWHPISVVKPYLTLLSECSNPDITLEGAGAL
|||
DB 796 TPQEQWDGVPGLPDCAPPKGIQMLWHPISVVKPYLTLLSECSNPDITLEGAGAL
|||
QY 876 GSWKSVYIRAAVRKEKGLPILVELLRINDRNVCAVATARNMALDVRNKELIG
|||
DB 856 SNWKEFAVIRGGRPKRKGILVELLRINDRNVCAVATARNMALDVRNKELIG
|||
QY 936 DLAVRLPGNNNSNTASKAMSDDTVTAVCTLHEVITKNMENAKALRDAGGIEKL
|||

ELLRRTGNAEIQKQLTGLLWNLSTDELKEELIADALPVLADRVIIIPFGWCDGNS 332
C1QLHSSQVLRNATGCLRNVSAGAEARRRMRECDGLTDALLVYVIGALGSSEIDS 758
EVV--DPEVFENATGCLRNLSA-DAGRTWRNYSGLIDSELMAYVQMCVAASRCD 389
VCVILRNLSYRLAAETSQCHMGTDDELGLLCEANGKDA---ESSGCGKGGKK 814
NCVILNLSYRLDAEVPTRYQ-----LEYNARNAYTEKSTGC-ESNKS 438
QDQWDGVLPLDCAEPKGIQMLWHPISIVKPYLTLLSECSNPDTLEGAAGALONLA 874
NNYD--CPLFEEINPKSGWLYSHDAIRTYLNLGKSKDATTLEACAGALQNTT 496
KWSYVIRAAV-RKEKGLPIVELLRINDNRVVCATARNMALDVNRKELIGKYA 933
LMSGMSQLGLKEKGLPOLARLLQSGNSDVRSGLLSNNMCKPILMKVMGNQV 556
VHLRPG--GNNSNNTASKMSDDTVTAVCTLHEVITKNMENAKALRDAGGIEKLV 991
TRLTSHGTGNTS-----SEDILSSACTVRLNLMASQPOLAKQVPSSSLNII 609
SKGDKSPKVVKAQSVLNSMWCYRDLRSYKKGWSQ 1033
3SA---SPKAAEAARLLSDMSKELQGLVROQGFDR 648

Location US/09041886
72
ION:
cedesen, Dale E.
abizadeh, Sharro
TION: Proapoptotic Peptides, Dependence
TION: Polypeptides and Methods of Use
JENCES: 72
3 ADDRESS:
Campbell & Flores LLP
70 La Jolla Village Drive, Suite 700
Diego
ifornia
ited States

ABLE FORM:
Floppy disk
IBM PC compatible
STEM: PC-DOS/MS-DOS
ntentin Release #1.0, Version #1.25
ATION DATA:
NUMBER: US/09/041.886

ION:
ON INFORMATION:
ell, Cathryn A.
NUMBER: 31,815
CKET NUMBER: P-LJ 2626
ION INFORMATION:
(619) 535-9001
(19) 535-8949
SEQ ID NO: 23:
ACTERISTICS:
5 amino acids
linear
protein

4.0%; Score 256.5; DB 3; Length 1185;
arity 20.8%; Pred. No. 8.6e-10;
nservative 87; Mismatches 268; Indels 329; Gaps 40;
NTSNGDGETTETTSAILASVKEQELQFRLTREAEARQIVASQLERCKLOSETG 90

Db 34 SPGVSTSSSDG-KAEKSRQTAKARVEEASTPKVKNQGRSE-EISESEETNAF
Qy 91 SMSMSAEBQFQWQSQDKDIEDETTGUELVDSICIRLSQESGILDPOYSTGE
Db 92 TEQELPFPQSPDLSDLG-----RSLNDDGSDPRIDQDN
Qy 148 PSLLSQSALQINS-----KPEGSGFQ-----
Db 133 PSYISPGSVENDSDSSGLSGQPARPYHPPLPPPPQPPDSTPRQPEASEFPHPS
Qy 171 SYHSNQ---TIALGETTFS-----QLPARGTOARATGQSFQ---GTTTRAGHLA
Db 193 GYHAPMEPTSRMFQAPPAPPHPOLYPGTGGVLSPGMPGKGGGAASVGGPN
Qy 219 APP-----PPPRFPFAPSLGSAFHLPDAPFAA-----A
Db 253 HPPPTTPISVSSGASGAPPTKPTTTPVGGG--NLPSAPPPANPPHVTNLPPLPPPA
Qy 253 YYSSTLP---APPRGG---SPLAAPQ--GGSPTKLORGSAPEGATYAAPRGSSP
Db 311 NNASAPPGIGAQLPGLPGLPSPYAMGQMGGLPPGPEKGTLPAPSHSLPPASSA
Qy 305 SRLAKSYSTSPI---NIYVSSAGLSPIRVTS-----PPTVOSTISSSPIH
Db 371 MRFPYSSSSSSAAASSSSSSSSSASPPASQALPSYHSPFPPTSLISVSNQPPK
Qy 352 TIGTYATLS-----PTKELVHAS-----
Db 431 SLFSAQVMSQGPPIPPPYGRLANSNAHPGFPPTSTGAQSTAHPPVSTHHHHQQQ
Qy 372 YSKHSQELYATATLQRPQS---LAAGSRASYSSQHG---LGPRLALQ---SP
Db 491 QQQQQQHHCNSGPPPPGAPHPLEGG-----SSHHAHPYAMSFSLGSLRPPPGP
Qy 419 --HIDIYEDRVYQKPPMRSLQ---SQGD-----PLPPAHT
Db 546 PPHSQVSYQAQGNPPVSSSSSSSTSGSYPCSHPSQGPQAGPYPPFPVPT
Qy 451 -----GTYRTSTAPS--SPGVDSVPL
Db 606 SATLSTVIATVASSPAGYKTASPPGPPPYKRAPSPGAYKTATPPGKPG--SPPS
Qy 474 SOHQPN-----AAATFORASYAAGPASNYADPYRLOQYCPVESPVSKSGPA---
Db 664 TPGYGTSPGAGFTKPGSPVGP-----GDLPPAGPSGLP
Qy 525 PEGTLARSP--SIDSTQKOPREFGWRDPEL-----PEVIQMLQHQFPVSQ
Db 706 PPAAPASGPPLSATQIKQEPAE-EYETPESPVPAPSPSPPPKVDV-----PSHA
Qy 572 AYLOHLCFGDNKIKAEIRROGGIQLLVLLDHEMTVHRSACGALRNLYVKANDD.
Db 760 RFNKHLDRGFN-----SCA--RSDLYFVPLEG.
Qy 632 LKNCGGIPALVRLLRKTTDLRE 655
Db 789 KKEAD---LVEKVRREAEQARE 808

RESULT 10
US-09-266-225D-18
; Sequence 18, Application US/09266225D
; Patent No. 6573364
; GENERAL INFORMATION:
; APPLICANT: Nandabalan, Krishan
; APPLICANT: Kingsmore, Stephen
; APPLICANT: Tchernev, Velizar
; TITLE OF INVENTION: Isolation and Characterization of Hermansky-Pud.
; TITLE OF INVENTION: Syndrome (HPS) Protein Complexes and HPS Prote.
; TITLE OF INVENTION: Interacting Proteins
; FILE REFERENCE: 15966-523
; CURRENT APPLICATION NUMBER: US/09/266,225D

us-09-501-171a-4.rapb

	Query Match	88.2%	Score 5625;	DB 13;	Length 1084;
	Best Local Similarity	99.7%;	Pred. No. 0;		
	Matches 1075;	Conservative	0;	Mismatches	3; Indels 0; G
QY	148	PSLLSQSALQINLKPEGSFOYPASYHSNQTALGETTTSQLPARGTQARATGQSFS			
Db	7	PSLLSQSALQINLKPEGSFOYPASYHSNQTALGETTTSQLPARGTQARATGQSFS			
QY	208	SRAGHLAGPPAPPPPPPPPPPPPPPSGLSGAFHLPDAPAAAAAALYYSSSTLPAPP			
Db	67	SRAGHLAGPPAPPPPPPPPPPPPPPPPSGLSGAFHLPDAPAAAAAALYYSSSTLPAPP			

Location US/10071900
US20020127541A1

TION:

George-Hyslop, Peter H.
mens, Johanna

aser, Paul E.

TION: Alzheimer's Related Proteins and Methods

TION: of Use

: 1034/1F810-US1

ATION NUMBER: US/10/071,900

DATE: 2002-02-08

ION NUMBER: US/09/227,725

ARE: 1999-01-08

ID NOS: 4

SEQ for Windows Version 3.0

o Sapien

41.4%; Score 2642; DB 13; Length 1193;

larity 47.1%; Pred. No. 6.6e-166;

Conservative 199; Mismatches 349; Indels 120; Gaps 39;

PDQSSASEKTSLSPLGINTNGDSETETT-SAILASVKEQELQFRLTRELEAER 73

PEQASLVEEQPQTRQEA-ASTGCMPEPETATTILASVKEQELQFRLTRELEVER 59

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

ASQLERCKLGSETSMSSMAEAFQFOWQOD-----GQKDIEDLTGTLGVDS 127

14:26:19 2004

us-09-501-171a-4.rapb

ON: Xaa equals any of the twenty naturally occurring L-amino acids
FEATURE
ON: Xaa equals any of the twenty naturally occurring L-amino acids
FEATURE
ON: Xaa equals any of the twenty naturally occurring L-amino acids
FEATURE
ON: Xaa equals any of the twenty naturally occurring L-amino acids
36.0%; Score 2296; DB 15; Length 785;
arity 58.8%; Pred. No. 3e-143;
nservative 117; Mismatches 173; Indels 34; Gaps 14;
PTVQSTISSPIHQLSSITGTATLSPT-KRLVHASEQYKSHSQELYATATLQRP 388
SOGQ-VGSSP--KSGMTAVPQHLGSLQRTVHMEQFGQQQYDIYERMPRP 64
AGRAYSQHGLGPELRALQSPHHIDPIYEDRVYKPPMRSLQSQGDPLPPA 448
GLRSSYASQHSQGLDRLSAVSPDLHITPIYEGRTYSPVYRSPNHTVE-LQGS 122
RTSTAPSPGVDSVPLORTGSHGQPNAAATFORASYAAGPASYADPYRLOY 508
RT-----GSGVIGN--LQRTSSOR-----STLYQRNNYALNTATYAEYRPIQY 171
SPYSKGPALPEGTIARSPDSISQKOPREFGWRDPPELPEVIOMLQHOFFPSVQS 568
CNYNRLQHAVPADGTTSPDSISQKOPREFAMRDPPELPEVIMLQHOFFPSVQA 230
LQHLCEGDKNIYABTRQGGIQLVLDLDRHMTVEHRSACGALRNLYVCKANDNN 628
LQHLCEGDKNVMKVCRLGGIKHLVDLLDRHVLVEQKNACGALRNLYVCKANDNN 290
QNGGGTALVRLKRTDLEIRELVTGLVNLSSCDALNMPITODALAVLTNAVII 688
QNGGGTALVRLKRTDLEIRELVTGLVNLSSCDALNMPITODALAVLTNAVII 350
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 748
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 410
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 808
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 469
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 865
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 529
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 925
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 589
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 985
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 644
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 1045
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 704
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 1103
QNSPLQDDRKIQHSSQVLRNATGCLRNVSAGAEARRRRECGLDALLXVIQ 759
SR 1109

Db 760 GDATHK 765

RESULT 6

US-10-205-823-80
; Sequence 80, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Wensey, Angela M.
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVEN
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80
; LENGTH: 941
; TYPE: PRP
; ORGANISM: Homo sapiens
US-10-205-823-80
Query Match 21.4%; Score 1362; DB 14; Length 941;
Best Local Similarity 33.1%; Pred. No. 3e-81;
Matches 357; Conservative 162; Mismatches 323; Indels 236; G

Qy 41 DGSSETTSAILASVKEQELQFERLRELAERQIVASQLERCKLG-SETGSMSSM
Db 2 DSEVESTASILASVKEQEAQFEKLTALREERHVSQAQLERVRVSPQDANPLMAN
Qy 100 EQFQWQSQDQKDEIDELTGLVDSQIRSLQESGILDPQDYSTGERPSILLSQSA
Db 62 RHQNGRFVGDADLEKQKPSDLKL-----NG---QDHS-----LYST
Qy 160 SKPEGSFYQASYHSNQTALGETTFSQLPARGTQAPATQCSFSGTTSRAGHLA
Db 103 QEPG-----QIVETYTEEDP-EGAMSVSVSVETSDDGTTTTRT-----
Qy 220 PPPPPPPPPFAPSLGSAFHLDPAPAAAAAALYYSSSTLPAPPPGSGPLAAPCGG
Db 138 -----ETIVKVKVATVTTTVPVAMGPDGL-----P
Qy 280 LQSGAPEGATYAAPRGSSPKOSPSRLAKSYSTSPINIWSSAGLSPIRVTSP
Db 138 -----ETIVKVKVATVTTTVPVAMGPDGL-----P
Qy 340 TISSSPHQLSS-----IGTVATLSPTKRLVHASEQYKSHSQELYATA
Db 169 SVNNYIQLGRFRKNGNGGPGVVGQAGTATLPRNFHYPPDGYSRHYEDGY---
Qy 388 PGSLAAGSRASYSSQCHLGP-----ELRALQSPHHIDPIYEDRVYKPPMRSLQ

-----SDNYGSLRVTRIEBRYRPSMEGYRAPSRQD-VYGPQPVVRVGGSSVD 271
PAHGTGTRTSTAPSPGVDSV-----PLQRTGSHQHPQNAATAATFORASYAAG 494
RFHPPEYGLEDDQDSMGVDDLDYGMMSDYGTAARTGTG----- 311
NYADPYQLQYCPVESPYKSGP-----ALPPEGTIARSPSIDSTOK-DPRE 544
--SPRRRLR---SYEDMIGEVSDQYVWAPLAQHERGSLA---SLDSURKGGPPP 360
RDPELPEVIQMLQHPFVSQNAAYLQHLGFGDNKIKAEIRROGGIQLIADLDHR 604
RQPELPEVIAMIGFELDAVKSNAAYLQHLGFCYRNDKVKTVRKLKGPVLVGLLDHP 420
VHRSACGALNRYLVKANDNDKIALKNCGGIPALVRLLRKTTDLREILVTVGLVNL 664
VHLGACGALKNISFGR-DQDNKIAIKNCDGVPALVRLLRKARDMDLTVITGLVNL 479
DALKMP1IQQDALAVLTNAVIIPHSWGNSPLQDDRK1QLHSSQVLRNATGLRNVS 724
DSIKMEIVDRHALHALTDEVIIIPHSWNEREPNEDCKPRHIEWESVLTITAGCLRNVS 539
EARRMRECDGLTDALLYIQSALGSSEIIDSKTVCNVCILRNLSYRLAAETSQGH 784
EARKLRCDGLVDALFIVQAEIGQKSDSKLVENCVCILRNLSYQVHREIPQ--- 596
DELGLLCEANGKADSSGCGWKKKKKKK--QDQWDGVGPLDPAEPKGIQMLW 842
ERYQEAAPNVANNTGPHAAACFGAKGKPKPIEDPANDTVDFPKRTSPARGVELLF 654
IVKPYLTLLGECSPDPTLEGAAGALQNLAAAGSKWWSVYIRAAVRKEKGLPLVELLR 902
VVRYYISLLKESKTPALIEASAGAIQNLGAGWTYGYIRSAIRQEKALSAIADLLT 714
DRVVCVATALRNMALDVNRKELIGKYAMRDLVHRLPFGNNSNNNTASKAMSDDTVTA 962
ERVVKAASGALNLAVDARNKELIGKHAIPNLVKNLPGQCN---SSWNFSEDTIVS 771
TLHEVITYKNENAKALDAGGIEKLVGISKGDKHPKVVKAAASQVLSNMWQYRDL 1022
TINEVIAENLEAAKKRETQIEKLVLINKS--GNRSEKEVRAAALVLOTWINGYKEL 829
YKKGWQYHVFVASSSTIETR-ORPYSSSRTPSISPVVSPNNRSASAPASPRE 1079
LEKEGKKSDFQVNLNNAASQSSHSYDDSTLPLIDRNQKSDNYS-----TENE 882

7
Application US/10094749
US20030219741A1

TION: TAKAO

GIYAMA, TOMOYASU

SUKI, TETSUJI

KAMATSU, AI

TO, HIROYUKI

HII, SHIZUKO

MAMOTO, JUN-ICHI

ONO, YUUKO

O, YURI

SUKA, KAORU

GAI, KEIICHI

IE, RYOTARO

MECHIKA, ICHIRO

KI, NAOHICO

SHIKAWA, TSUTOMU

SUKA, MOTYUKI

GAHARI, KENJI

SUHO, YASUHIKO

TION: NOVEL FULL-LENGTH cDNA

: 084335/0160

ATION NUMBER: US/10/094,749

CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
PRIOR FILING DATE: 2001-09-14
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2227
LENGTH: 834
TYPE: PRT
ORGANISM: Homo sapiens
US-10-094-749-2227

Query Match 20.1%; Score 1282.5; DB 15; Length 834;
Best Local Similarity 45.1%; Pred. No. 4,76e-76;
Matches 321; Conservative 94; Mismatches 220; Indels 77;

QY 13 GAMPVDPQSSASEKTSLSPLNTSGDSETETT-SAILASVKBOELQFRLT
DB 130 GMPAPEQASLVEEGQFQTRQEA-ASTGFMPEPTATTILASVKEQELQFRLT
QY 72 ERQIVASQLERCKLSETGSMSSMAERQFQWQSQD-----GQKDI EDELTTG
DB 189 ERQIVASQLERCKLGAESPASIASTSTSKSPWRSTDVENTGSKPRVSDAVQPN
QY 126 SCIRSLQESGIL-DPDQYSTGERPSSLQSALQNLKSPGSGFOYPASYHSNQTLA
DB 247 --IRTEPEQGTLYSPQTSLHERSLGNSRSSTOMNSYSDSGYQEAQSFHNSQNV
QY 180 ----LGETTPTSQLPARGTQARATGQSFQ-GTTSRAGHLAGEPAPAPPPP-----
DB 305 QHSHFSGTNNHV----RNSRAEGQTLVQPSVANRAMRVSVSPRAQSPSVI
QY 225 PREPAPSLGSAFHLP---DAPAAAAALYYSSSTLPAPPRGSGSPLAAPQGS
DB 361 PGRSLRTSLGSGFGSPVTDPRPLNPSA---YSSSTLPA-ARAASPY-SQRPAS
QY 282 RGSAGEGATYAAPRGSSPK-QSPSLAKSYSTSSPINIVVSSAGLSPIRVTSPE
DB 416 RIGSVTSROT-SNPNGPTFOYTARV-----GSPILTIDAQT----RVASPS
QY 341 ISSPSIHLQSSITGTATLSPT-KRLVHASEQYKSHSOELYATATILQRPGLAAG
DB 463 GSSSP--KRSGMTAVPQHLGSLQRTVHDMEQFGQQQYDIYERWVPRPDSL-TG
QY 400 SSOHGLGPELRALQSPHHIDPIYEDRVYQKPPMRSLSQSQGDPLPPAHTGYR
DB 520 ASQHSQLGQDLRSVSPDLHITPIYEGRTYSPVYRSPNHGTVE-LQGSQTALYR
QY 460 SSPGVDSVPLQRTGSHQHPQNAATAATFORASYAAGPASVADPYQLQYCPVES
DB 575 GSVIGN--LQRTSSQ-----STLTYQRNNYALNTATYAEYRPIQTRVQ-EC
QY 520 GPALPPEGTIARSPSIDSIQKDPREFGWRDPDELPEVIQMLQHPFVSQNAAYL
DB 627 QHVPADDGTTSPSIDSIQKDPREFAWRDPDELPEVIHMLQHPFVSQNAAYL
QY 580 GONKIAETRRGGIQLIADLDHRMTEVHRACGALRNLYVTKANDNDKIALKN
DB 687 GONKIVMEVCRLLGGIKHLVDLDRHVLQVKNACGALRNLYVFGKSTDENKIAMKN
QY 640 ALVRLRKTTDLREILVTVGLNLSNCCDALKPIIQQDALAVLTNAVIIPHS 69
DB 747 ALVRLRKSIDAEVRELVTGVLNLSNCCDAVKGTTIIRDALST-THSVLWKHN 79

RESULT 8

US-09-925-301-1274

; Sequence 1274, Application US/09925301

; Patent No. US20020052308A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

PA106
 TION NUMBER: US/09/925,301
 DATE: 2001-08-10
 ON NUMBER: PCT/US00/05882
 TE: 2000-03-08
 ON NUMBER: 60/124,270
 TE: 1999-03-12
 D NOS: 1694
 tIn Ver. 2.0

sapiens

ION: Xaa equals any of the naturally occurring L-amino acids

arity 10.7%; Score 684; DB 9; Length 425;

conservative 76; Mismatches 144; Indels 20; Gaps 7;

EEARRRRECDGLDALLYVQSALGSEIDSKTVENCVCILRLNLSRLAAETSGQ 782

SEARRKRECDGLDALIFIVQABIGQKDSKLVENCVCILRLNLSYQVHREIPQ- 59

TDELDGLCGGANGKAESSGCWKKKKKKK--QDQDGVGVPDCAEPKGIOM 840

AERYQEAAPNVANNTGPHASCFAGKKGKPIEDPANDTVDFPKTSPARGVEL 115

SIVKPYTLSECSNPDTLEGAAGALONLAGSKWSYVYIRAAVRKGLPILVEL 900

EVVRIYISLLKESKTPAILLEASAGAIQNLACGRWYGYKIRSAURQKALSIAIDL 175

NDRVVCATATRLNMLDVNKLIGKVMRDVHRLPGGNNNSNTASKMSDDTV 960

HERVVKAASGALRLNLDVARNKELIGHAIPNLVKNLPGSQON---SSWPFSEDTV 232

CTLHEVITXNMENAKALRDAGGIEKLVGSKSGDKHSPKVKAASOVLSNMQYR 1020

NTINEVTAENLEAKKRETCGIEKVLINKS--GNRSEKEVRANALVLQTIWYK 290

LYKDGQSGVHFVASSSTIERDR-QRPYSSRTSPISFVRVSPNNRSASAPSPRE 1079

PLEKEGKKSDFQVNLNNAKSQSSHSYDSTLPLI-----DRNQKSKKEP--DRE 343

KERTDYECTGSGNATYHCAKEHTSRKD 1111

SNMGSTKSLDNNYSTPNERGDHNRITLD 375

Application US/09864761

20048763A1

ION:

Sharon G.

David R.

zel, David K.

Wensheng

ION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

ION: GENE EXPRESSION ANALYSIS BY MICROARRAY

Acemica-X-1

ION NUMBER: US/09/864,761

DATE: 2001-05-23

ON NUMBER: US 60/180,312

TE: 2000-02-04

ON NUMBER: US 60/207,456

TE: 2000-05-26

ON NUMBER: US 09/632,366

TE: 2000-08-03

ON NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 39492
 ; LENGTH: 86
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC005610.1
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.5
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.3
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.3
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.5
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.2
 ; OTHER INFORMATION: SWISSPROT HIT: Q10145, EVALUATION 1.60e+00
 ; US-09-864-761-39492

Query Match 7.4%; Score 475; DB 9; Length 86;
 Best Local Similarity 100.0%; Pred. No. 5.5e-24;
 Matches 86; Conservative 0; Mismatches 0; Indels 0; G

QY 1140 VSAQPVQEPSPKDYETYPQFQNSRNYDESFEDQVHHRPPASEYTMHLGLKSTG
 Db 1 VSAQPVQEPSPKDYETYPQFQNSRNYDESFEDQVHHRPPASEYTMHLGLKSTG
 QY 1200 FYSAARPYSELNYETSHYPASPDWV 1225
 Db 61 FYSAARPYSELNYETSHYPASPDWV 86

RESULT 10
 US-10-094-749-1831
 ; Sequence 1831, Application US/10094749
 ; Publication No. US20030219741A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ISOGAI, TAKAO
 ; APPLICANT: SUGIYAMA, TOMOYASU
 ; APPLICANT: OTSUKI, TETSUJI
 ; APPLICANT: WAKAMATSU, AI
 ; APPLICANT: SATO, HIROYUKI
 ; APPLICANT: ISHII, SHIZUKO
 ; APPLICANT: YAMAMOTO, JUN-ICHI

OTHER INFORMATION: MAP TO AC004648.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.3

ON: EXPRESSED IN ADULT LIVER, SIGNAL = 5.5
ON: EXPRESSED IN PLACENTA, SIGNAL = 5.3
ON: EXPRESSED IN HELA, SIGNAL = 4.8
ON: EXPRESSED IN HEART, SIGNAL = 4.9
ON: EXPRESSED IN FETAL LIVER, SIGNAL = 4.5
ON: EXPRESSED IN BONE MARROW, SIGNAL = 4.1
ON: SWISSPROT HIT: P30999, EVALUE 3.00e-13
ON: EST_HUMAN HIT: A1936774.1, EVALUE 1.00e-22
; 4.4%; Score 280; DB 9; Length 55;
urity 100.0%; Pred.No. 2.4e-11;
nservative 0; Mismatches 0; Indels 0; Gaps 0;
; 658
; 55

Application US/09864761
0048763AL

ION:

Sharron G.

el, David R.

el, David K.

l. Wensheng

ON: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

ION: GENE EXPRESSION ANALYSIS BY MICROARRAY

Aeomica-X-1

ION NUMBER: US/09/864,761

DATE: 2001-05-23

ON NUMBER: US 60/180,312

TE: 2000-02-04

ON NUMBER: US 60/207,456

TE: 2000-05-26

ON NUMBER: US 09/632,366

TE: 2000-08-03

ON NUMBER: GB 24263.6

TE: 2000-10-04

ON NUMBER: US 60/236,359

TE: 2000-09-27

ON NUMBER: PCT/US01/00666

TE: 2001-01-30

ON NUMBER: PCT/US01/00667

TE: 2001-01-30

ON NUMBER: PCT/US01/00664

TE: 2001-01-30

ON NUMBER: PCT/US01/00669

TE: 2001-01-30

ON NUMBER: PCT/US01/00665

TE: 2001-01-30

ON NUMBER: PCT/US01/00668

TE: 2001-01-30

ON NUMBER: PCT/US01/00663

TE: 2001-01-30

ON NUMBER: PCT/US01/00662

TE: 2001-01-30

ON NUMBER: PCT/US01/00661

TE: 2001-01-30

ON NUMBER: PCT/US01/00670

TE: 2001-01-30

ON NUMBER: US 60/234,687

TE: 2000-09-21

ON NUMBER: US 09/608,408

TE: 2000-06-30

ON NUMBER: US 09/774,203

TE: 2001-01-29

ON NOS: 49117

ax Sequence Listing Engine vers. 1.1

TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005610.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3
OTHER INFORMATION: SWISSPROT HIT: P30999, EVALUE 4.00e-08
OTHER INFORMATION: EST_HUMAN HIT: A126185.1, EVALUE 5.00e-17
US-09-864-761-38977

Query Match 3.9%; Score 246; DB 9; Length 50;
Best Local Similarity 100.0%; Pred.No. 3.7e-09;
Matches 50; Conservative 0; Mismatches 0; Indels 0; G

QY 880 WSVYIRAAVRKEKGLPILVELLRIDNDRVVCAVATLRNNALDVRNKELI 929

Db 1 WSVYIRAAVRKEKGLPILVELLRIDNDRVVCAVATLRNNALDVRNKELI 50

RESULT 14

US-09-864-761-39938

Sequence 39938 Application US/09864761

Patent No. US20020048763AL

GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID P

FILE REFERENCE: Aeomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

ION NUMBER: US 09/774,203

ATE: 2001-01-29

ID NOS: 49117

max Sequence Listing Engine vers. 1.1

o sapiens

TIION: MAP TO AC003089.1

TIION: EXPRESSED IN BRAIN, SIGNAL = 1.8

TIION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5

TIION: EXPRESSED IN LUNG, SIGNAL = 1.4

TIION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4

TIION: EXPRESSED IN PLACENTA, SIGNAL = 1.7

TIION: EXPRESSED IN HEART, SIGNAL = 1.4

TIION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4

TIION: SWISSPROT HIT: P30999, EVALUE 1.00e-09

TIION: EST_HUMAN HIT: BE177290.1, EVALUE 8.00e-18

38

larity 100.0%; Score 239; DB 9; Length 44;

Conservative 0; Mismatches 0; Indels 0; Gaps 0;

WRDELPDEVIMLOHQFSPVQSNAAAYLQHLCFGDNKIKAE 587

WRDELPDEVIMLOHQFSPVQSNAAAYLQHLCFGDNKIKAE 44

8 Application US/10128714

US20030119013A1

TIION:

ng, Bo

Wenqi

Shkoff, Daniel

udio, Carlos

oshkin, Alexey M

mieux, Sebastien M

TIION: Identification of Essential Genes in Aspergillus fumigatus and

TIION: Methods of Use

: 10182-018-999

ATION NUMBER: US/10128,714

DATE: 2002-04-23

ION NUMBER: US 60/285,697

ATE: 2001-04-23

ION NUMBER: US 60/287,066

ATE: 2001-04-27

ION NUMBER: US 60/295,890

ATE: 2001-06-05

ION NUMBER: US 60/303,899

ATE: 2001-07-09

ION NUMBER: US 60/316,362

ATE: 2001-08-31

ID NOS: 8603

ntin version 3.1

ergillus fumigatus

8

larity 3.6%; Score 227.5; DB 14; Length 1832;

Conservative 20.5%; Pred. No. 1.2e-05;

Conservative 181; Mismatches 503; Indels 485; Gaps 79;

SEKTSLSPLGNTNGDSETTTSAILASVKEQELQELTRLEABRQIVASQLE 81

QKRTSAEGFPPIITKRTSQVIDSINAVSDSMFSKQLP-----AAAEIN 319

KLGETGSMSSNAEE-QFOWQSQ-----DGQKDIEDELT--GLELVDSCIRSLQ 132

Db 320 NSLAVPTADGEIKNSPSEEDLAARWQAEISDDADETMTPTEDDILAAWQAEILD--
QY ATE: 2001-01-29
ID NOS: 49117
max Sequence Listing Engine vers. 1.1
o sapiens
TIION: MAP TO AC003089.1
TIION: EXPRESSED IN BRAIN, SIGNAL = 1.8
TIION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
TIION: EXPRESSED IN LUNG, SIGNAL = 1.4
TIION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
TIION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
TIION: EXPRESSED IN HEART, SIGNAL = 1.4
TIION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
TIION: SWISSPROT HIT: P30999, EVALUE 1.00e-09
TIION: EST_HUMAN HIT: BE177290.1, EVALUE 8.00e-18
38
larity 100.0%; Score 239; DB 9; Length 44;
Conservative 0; Mismatches 0; Indels 0; Gaps 0;
WRDELPDEVIMLOHQFSPVQSNAAAYLQHLCFGDNKIKAE 587
WRDELPDEVIMLOHQFSPVQSNAAAYLQHLCFGDNKIKAE 44
8 Application US/10128714
US20030119013A1
TIION:
ng, Bo
Wenqi
Shkoff, Daniel
udio, Carlos
oshkin, Alexey M
mieux, Sebastien M
TIION: Identification of Essential Genes in Aspergillus fumigatus and
TIION: Methods of Use
: 10182-018-999
ATION NUMBER: US/10128,714
DATE: 2002-04-23
ION NUMBER: US 60/285,697
ATE: 2001-04-23
ION NUMBER: US 60/287,066
ATE: 2001-04-27
ION NUMBER: US 60/295,890
ATE: 2001-06-05
ION NUMBER: US 60/303,899
ATE: 2001-07-09
ION NUMBER: US 60/316,362
ATE: 2001-08-31
ID NOS: 8603
ntin version 3.1
ergillus fumigatus
8
3.6%; Score 227.5; DB 14; Length 1832;
20.5%; Pred. No. 1.2e-05;
Conservative 181; Mismatches 503; Indels 485; Gaps 79;
SEKTSLSPLGNTNGDSETTTSAILASVKEQELQELTRLEABRQIVASQLE 81
QKRTSAEGFPPIITKRTSQVIDSINAVSDSMFSKQLP-----AAAEIN 319
KLGETGSMSSNAEE-QFOWQSQ-----DGQKDIEDELT--GLELVDSCIRSLQ 132

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

in search, using sw model

sil 22, 2004, 12:10:45 ; Search time 27 Seconds
(without alignments)

4364.245 Million cell updates/sec

-09-501-171A-4

77

MFARKPFGAALPLGAMPVDPQ.....PYSELNYETHYPASPDWV 1225

DSUM62

pop 10.0 , Gapext 0.5

3366 seqs, 96191526 residues

ts satisfying chosen parameters: 283366

3th: 0

3th: 2000000000

inimum Match 0%

aximum Match 100%

isting first 45 summaries

IR_78:*

pir1:*

pir2:*

pir3:*

pir4:*

the number of results predicted by chance to have a
c than or equal to the score of the result being printed,
ed by analysis of the total score distribution.

SUMMARIES

arch	Length	DB	ID	Description
5.1	1247	2	T42209	neural plakophilin
1.6	911	2	S28498	gene p120 protein
4.8	1050	2	T26395	hypothetical prote
1.2	725	2	S60712	band-6-protein - h
7.9	295	2	S60711	band-6-protein - b
3.9	1184	2	S50832	atrophin-1 - human
3.6	1006	2	T42731	atrophin-1 related
3.5	3938	2	T42761	Bassoon protein -
3.4	528	2	T47141	Gastric mucin (clo
3.4	1357	2	T29285	hypothetical prote
3.3	1560	2	T00080	hypothetical prote
3.3	2232	2	T34434	hypothetical prote
3.3	1198	2	T49726	hypothetical prote
3.2	806	2	JC4835	beta-catenin - Hyd
3.2	2649	2	T51023	hypothetical prote
3.2	1733	2	S27939	tensin - chicken
3.2	721	2	E70766	hypothetical prote
3.2	1487	1	EDBEF6	155K transcription
3.2	817	2	S51342	verprolin - yeast
3.2	1320	2	JC5630	TCOF1 protein - mo
3.1	2845	2	I49505	adenomatous polypo
3.1	839	2	T04859	extensin homolog F
3.1	580	2	T43481	probable mucin DKF
3.1	1151	2	T18535	high molecular mas
3.1	954	2	T19765	hypothetical prote
3.1	1851	2	T19964	hypothetical prote
3.1	1211	2	T42230	AF4 protein - mous
3.1	2715	2	T13049	eyelid - fruit fly

30	194	3.0	660	1	QBEB3	BHLF1 prote
31	194	3.0	2843	1	RBUHAP	adenomatous
32	193.5	3.0	1032	2	T34433	hypothetica
33	193	3.0	1323	2	T30253	snalt prote
34	192.5	3.0	3942	2	T42730	Bassoon prc
35	191.5	3.0	781	2	A38573	beta-cateni
36	191.5	3.0	1896	1	RNFF2L	DNA-directe
37	190.5	3.0	1142	2	T00022	B120 protei
38	190	3.0	878	2	T21621	hypothetica
39	190	3.0	1487	1	EDBE1	immediate-e
40	189.5	3.0	781	2	S35091	beta-cateni
41	189.5	3.0	1099	2	A56155	tumor suppr
42	188.5	3.0	1148	2	F86403	probable tr
43	188.5	3.0	1273	2	S58782	SEC31 prote
44	188	2.9	1217	2	T42625	AF-4 protei
45	188	2.9	1460	1	EDBEIF	immediate-e

ALIGNMENTS

RESULT 1

T42209

neural plakophilin related arm-repeat protein NPRAP - mouse

N;Alternate names: plakoglobin/armadillo protein

C;Species: Mus musculus (house mouse)

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec

C;Accession: T42209

R;Paffenholz, R.; Franke, W.W.

Differentiation 61, 293-304, 1997

A;Title: Identification and localization of a neurally expressed membe

A;Reference number: Z22075; MUID:98002299; PMID:9342840

A;Accession: T42209

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1247 <PAF>

A;Cross-references: EMBL:U90331; NID:g2580536; PID:g2580537; PIDN:AA8

A;Experimental source: brain

C;Genetics:

A;Gene: NPRAP

A;Note: predominantly, if not exclusively, expressed in neural and neu

Query Match	96.1%;	Score 6129;	DB 2;	Length 1247;
Best Local Similarity	94.8%;	Pred. No. 1.3e-312;		
Matches 1185;	Conservative	8;	Mismatches 29;	Indels 28;
Qy	1	MEARKPFGAALPLGAMPVDPQPSASSEKTSLSGLNTSNGDGSSETTTSALLASVK		
Db	1	MEARKQSGAAPPFGAMPVDPQPSASEKNSSLSGLNTSNGDGSSETTTSALLASVK		
Qy	61	QFERLTRELEBAERQIVASQLERCKLGSETGSMSSMSAEQFQWQSQDQKDIIDE		
Db	61	QFERLTRELEBAERQIVASQLERCKLGSETGSMSSMSAEQFQWQSQDQKDIIDE		
Qy	121	LELVDSICRSLQSGIILDPODYSTERPSLLSOSALQNSKPEGSFOYPASYHSNQ		
Db	121	LELVDSICRSLQSGIILDPODYSTERPSLLSOSALQNSKPEGSFOYPASYHSNQ		
Qy	181	GEITPSQLPARGTCARATGOSFSGCTTSRAGHLAGPEPAPPPPPPPPPPPPPSGLS		
Db	181	GDTPASQLPARSTQARAAGSFSGCTTSRAGHLAGSEPA-PPPPPPPPPPPPPPSGLS		
Qy	241	PDAPPAALAAALYSSSTLPAPPRGGSPLAAPQGSPTKLQSGSAPGATYAAPR		
Db	240	PDAPP--AAALYSSSTLPAPPRGGSPLTTQGSPTKLQSGSAPGAAAYAAPR		
Qy	301	KQPSRLAKSYSTSSPINIVVSSAGLSPIRVTSPPTVQSTISSPHQLSTIGTY		
Db	298	KQPSRLAKSYSTSSPINIVVSSAGLSPIRVTSPPTVQSTISSPHQLSTIGTY		
Qy	361	PTKRLVHASQYKHSQELVATATLQPGSLAAGSRASYSSQHGHLPELRLQSP		
Db	358	PTKRLVHASQYKHSQELVATATLQPGSLAAGSRASYSSQHGHLPELRLQSP		

[illegible]

Query Match	Best Local Similarity	4.0%; Score 256; DB 2; Length 1184;
Matches 169; Conservative 78; Mismatches 250; Indels 294;		
QY 32 SP-GLNTNGDGETTETTSAILASVKQEOLOFERLRELEAEROIVASQERCKI	DB 34 SPGGVSTSSDD-KAEKSRQAKARVEEASTPKVKGGRSE-EISESEETVA	QY 91 SMSMSSSABEQFQWQSQDQXKDIEDLTTGLEVDSCIRSLQESGILDPDYSTK
DB 92 TEQELPRQPSDLDLSLDG	DB 93 PLSLSQSLQNLNS	DB 94 SPGGVSTSSDD-KAEKSRQAKARVEEASTPKVKGGRSE-EISESEETVA
QY 148 PLSLSQSLQNLNS	DB 133 PSYVPGSVENDSSSSGLSQPARPYHPPPLFPSPQDPDSTBRQPEASFEPIH	QY 171 SVHSNQ---TLALGETTSS---OLPARTQARATQGSFSQ---GTTSRAGHI
DB 193 GHVAMEPPTSRMFQAPPAPHPHPQLPYGGTGVLGSPPMGPKGGGAASVSGFE	QY 219 APPP-----PPPPREPAPSLGSAFHLDPAPAA-----	DB 253 HPPPTPTISVSSSGASGAPPTKPTTPVGGG--NLPSAPPANPPHVTNLPPLPPE
QY 253 YVSSSTLP---APRGG---SPLAAPQ---CGSPYKLOGGASPEGATYAAPRGSE	DB 311 NNASASPGLGAQLPGLHPLSPYAMGQGMGLPGPEKGTLPAPSPHSLPAPASSE	QY 305 SRLAKSYSTSSPI---NIVVSSAGLSPKRVTS-----PPTVQSTTSSSPI
DB 371 MRFPYSSSSSSSSAAASSSSSSSSSSAGFFPQASQALPSYPHSFPPPTSLVSQNPDE	QY 352 TIGYATLS-----PTKRLVHAS-----	DB 431 SLPSQAVMSQGPFPFPYGRLLANSNAHPGFPFPSTGAQSTAHPVSTHHHHHQQ
QY 372 YSKHSQELYATATLQRGS---LAAGSRASYSSQHG---LGPRLALQ-----S	DB 491 QQQQQQHHGNSGPPPPGAPFHPLEGG-----SHHAHPYAMSPSLGSLRPPYPPG	QY 419 --HIDPIEDRVYQKPMRSLSQ-----SQGD-----
DB 546 PPHSQVSYSQAGPNGPVPVSSSSSSSSSTSGSYPCSHSPSQGPGAGPYFPFPPVP	QY 451 -----GTYRTSTAPS-SFGVDSPV	DB 606 SATLSTVIATVASSPAGYKTASPFPFPYGRAPSPCAVKTATPPGYKFG---SPP
QY 474 SQHPQN-----AAAATFORASYAAGPASNVADPYRQLQVCPSPVESPVYKSGPA--	DB 564 TGVGVGTSRNGCTKVGCSNVVVB-----	

```

LARSP--SIDSIQKDPREFGWRDEL-----PEVIOMLQHQPFSVOSNAA 578
      |||::|||::|||::|||::|||::|||::|||::|||::|||::
PASGPPLSATQIKOEPAE-EYETPESVPFARSPPPPKVDV-----PSHASQSA 759
HLCFGDN 582
HLDRGFN 770

piens (man)
#sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
3

MBL Data Library, March 1995
3 : G08343
RNA : translated from GR/EMBL/DDBJ
<MAR>
: EMBL:U23851; NID:g915325; PID:g915326

B37 : GDB:270336; OMIM:125370
p-12p

          3.9%; Score 250; DB 2; Length 1184;
arity 20.9%; Pred.No.3e+05;
conservative 84; Mismatches 260; Indels 346; Gaps 42

LNTNSNGDGETTETSALLASVKQEQLQFERLTRELEAEQIIVASQLERCKLGSETG 90
||||::|||::|||::|||::|||::|||::|||::|||::|||::
VSISSSDCK-----AQRQTAKARVEASTPVKNKO-----G 71

MSSAREQFQMWSQDGKDIEDEL-----TTGLELVDSICIRSLQESGIILPDQDYSTG 145
||||::|||::|||::|||::|||::|||::|||::|||::|||::
ISESEB---ETWAPKKTKTEELPRQSPSDDLSDLG--RSLNDGGSDRPDIQDD 126

-PSSLQSALQINS-----KPEGSGFQ----- 167
SPIYSFGSVENDSDSSGLSQGARPYHPFLFPSPQPPDTTPRQPEASFEPHP 186
ASHVSNQ----TLALGETTFS-----QLPARGTQARATGSFSQ--GTTSRAGHL 213
TGYPHAPMEPTSRMFQAFCAPPHPHQLPQGTGGVLGFFPMGPKGGGAASVGPF 246
PAPPP-----PPPPEPPAPLSGAFHLPDAPPAANAAAAYIGSSTLP 260
|||::|||::|||::|||::|||::|||::|||::|||::|||::
QHPEPTTPISVSSSGASGAPTTPPTTVPGVG--NLPGAAPP--PANPHVTNPFL 301
-----GGSP-----LAAPQ-----GSPTKLORGSAPEGATYAAPR 296
LRLENNASAPPGLGAOPLPHGLPSHAMQGIGLLPGKEKPTLASPSHLPPA 361
KQSPRLAKSYSTSSI-----NIWSSAGLSPIRVTS-----PPTVQSTISS 343
|||::|||::|||::|||::|||::|||::|||::|||::|||::
PAPMPRFYSSSSSSAAAAASSSSSSSSSSASPASQALPSYPHFPPPTSLSVSN 421
QLSSTTGTVATLS-----PYKLHVHAS----- 369
|||::|||::|||::|||::|||::|||::|||::|||::|||::
YTQPSLPQAVMSQQGPPPPPPYGRLLANSNAHPFPFPPTGAOSTAHPPVSTRHHH 481
--EQYKSHSOELYATATORPGS-----LAAGRSASYSSOHGH--LGPLRALQ-- 414
|||::|||::|||::|||::|||::|||::|||::|||::|||::
QQCQQCQQCQQHNGSGPPPPPCAFPHPLEGG-----SSHHAIFAMYSLSGLRAPY 536
EH-----HIDIYEDRVYQKPMRKSLSQ-----SQGD-----FLP 446
AHLPPHSQVSYSEQANGFPPVSSSSNSSSSSTSOGSYPCSHPSPOGQPGAPYDFP 596

```

```

-----SGSPFFSNMNLPPPPALKPLSLSTHPP-----SAHPPPLQLMP 430
XGDLPLPAHTGTVRTSTA-----PSSPGVDSVPLQRTGSH-----GPQNAAAAT 485
PPLSPPAQPPGLTQSQSLPPPAASHPTTGLHQVFSQSPFPPQHPVPVPGPPPTPPS 490
TASY-AAGPASNYADPYRQLOQCFSPESYSK--SGPALFPBEGTLARSPSIDSIQK-- 540
TSTTPAGPSSSSQPP-----CSAAVSSGNGVPGAPSCPLPAVQIKKEALDEAEPE 544
--DPRFEGWRDPELPEVIQMLQHQFFSVQSNAAAYLOHLCFGDN 582
PPEPRS---PSPE-PTWDT-----ESHASQSARFYKHLDRGYN 581

. rat
; brain-specific synapse-associated protein
; norvegicus (Norway rat)
100 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
61
; Marti-Vila, L.; Langnaese, K.; Richter, K.; Kindler, S.; Soyke, A.; Wex,
; 499-509, 1998
; a novel zinc-finger CAG/Glutamine-repeat protein selectively localized
; at: Z22249; MUID:98345363; PMID:9679147
61
; nary; translated from GB/EMBL/DBJ
; MENA
; 8 <DIE>
; : EMBL:Y16563; NID:g3413503; PIDN:CAA76287.1; PID:g3413504
; urce: strain Sprague Dawley; brain

y be involved in cytomatrix organization at the site of neurotransmitter
of the presynaptic cytoskeleton
d coil; zinc finger

3.5%; Score 224.5; DB 2; Length 3938;
arity 19.5%; Pred. No. 0.0034;
Conservative 177; Mismatches 632; Indels 483; Gaps 67;

TPGA-APIGAMPVPPQPSASEKTSLS--PGLNNTSGDS----- 43
TPSALAGGQLFVAGAAAGPPTPGLGLVPGFGPGFPGSVSRRLDPKEPLGSQA 88
---ETETTSAILASVKE---QELQFERLTRELAERO----- 74
TPKQASATAPGRESPRETRAQGLSQEAGRPRTLOVDSRTQSRGSPSVSPDRGST 148
-----IVASQLERCKLGETSGMSMSAEQFQWQDQK-----DIED 115
PYSVPQIAFLPSSTL--CPI-CKTSDLTSTG--QPNFNTCTQCHNKVCNQCQGFNDP 204
TGLELVDSICRSLOESGILDPODYTGEPRLSLSQALQLSKPGSGFYQPSYHSN 175
QVKEWL--CLNCQMORAL--GMDMTTAPR---SKSQQLHS--PALS---PAHSAPK 252
ALGETTPQLPARGTQARATGQSFQSGTTSRAGHLAGEPAPPPP-----PP-- 225
CKPEQERSRSPG-----ATQSGPQAEARATSVPGTQATAPPEVGRVSPQPLS 306
-----PREPAPSLG-SAFHLPDAPPA-----AAALYYSSSTI----- 259
STAEPFRPAGEAQGKSATTPVPSGLGAAEQTGGLTGKFLGGLGASLLTQASTLMSYQP 366
-----PAPRPGSGPLA-----APQGSPTKLQSGSAPGATYAAPRGSSPKQ 302
TGQGPSKGPPIKVFSDASKEAGRPFGSGGP-----GPTGAKTEFGPR-TGPGS 421
RLAKSYSTSSPINIVVSSAGLSPIRVTPPTVQSTISSPIHQLSSTIG----- 354
IALAKTGTPSPKHGRADHQAASK-AAAPKMTMFERAAACPLCAELNVGSRGPANYN 480

```

```

355 -----TYATLSPTKRLVHASEQYSKHSQ--ELYATATLQRPQSL-----
481 TCITACKLRVCTLCGFNTPTPLHVEKTEWMLCLNCQTKRLLEGSLGEPAPLPLTPQ
392 ---AAGSRASYSSQHGHLGPELRALQSPHEHIDIYEDRVYQKPPMRSLSSQSGI
541 PQRAAG--ASPLKQKGPQGP-----GQPSGSLPPKASPOAKAKASPOAKASPOA
446 -----PAAHTG--TVRTSTAPSSPGVDSVPLQRTGSHGPNQAAAAAATPQRA
594 EPSKTSASAPEKKTGIPVKAEPVPKPPETAVPPGCTPKAKSGVKTDPATPVVKI
496 ASNYAD-----PYRQLOQCFSPESYSKGPALPPEGTTLARSPSIDSIQKPRE
654 KSGAEPEVPKPYSQ-----DLRSFQSLSDTGYSSDGVSSSQSEITGVVQOQVEI
545 -FGWRDPELPEVIQMLQHQFFSVQSNAAAYLOHLCFGD-----NKKIABIRI
710 VTGPRPSPSELHKVGSMPRLPEAQAVA-----PSGEWSKPPSGSAVEDQKRRI
597 LVDLLDHR-----MTSVHRSACGALNLV--YKANDDNKIALKNC-----G
765 MPEAFDDEELGDILEDDSLAMGRQREQQDTAESDDFGSQLRHDYVEDSEGA
643 RLARKTTDLREIRLVTVGLMNLSSCDALKMPIIQDALAVLTNAV-----IIPHSGI
825 PQPPARADMTDEEPMRQI-----LEMSABEDNLEEDDTAVSGRLAKHGA
597 --PLQDRKTLQHSQVLRNAT-----GCLRNVSSAGEI
877 PRPESSQESVALPKRRPLPHNATTGYELLSEEGPAETPDGALQGLRRFKTIGLI
734 -----RECGLTDLALVYIQSALGSSSEIDSKTVENCVCILRNLA
937 WTSLSLDLGQSDPNLDREPE-----LEMESLTGSPEDRSRGHSHSTLPASTP
774 -----RLAAETTSQOQHMT-----DEL-----
990 SPTSLSLEEDSDSPSRQRLBEAKQORXARHRSHGFPLLPTIEDSSEEEELREI
796 GEANGKDAESSGCGWKKKKKKKSQDQW-----DGVGPLPDCAPPPK
1050 EQEKXREVEQOIRNSTARKTRRDKELRAQRRRRRSKTPPSNLSPIED-ASPTFI
842 ---WHPISVKPYLTLLSECSNPDTLEGAALQNLAAAGSMKSVYTRAAVKEI
1109 EMELHSSCSCEYSFSPSLDSEAEITLG--GPTLYKSGS-----E
898 VEL-----LRIDNRVVCATAARNMALDVNRKELIGKYAMRDLVHRI
1154 MSLCSPTETPSSGSTTTPSSGRPLKSAEEAYEDM---MRKAEILLQRCQQAAGARI
947 SNNT-----ASKAMSDDTTAVACCTLHEVITKNMENAKI
1211 SQTPGRSQSGFEYQDTLDHDYGGRAQPAADGTPAGLGATVYEEIILQTSOIAI
986 GIEXLVGISKSGDKHSPKVVVKAASQVLSNMWQYRDLRLSLYKKDQWGSQYHFVAS
1270 -----SSRDLATFEDKKKEQFLNAESAAYMD---PMKQNGGP---LTPG
1046 DRQPYSSRTPSISPVVSPNNRSA-----SAPASPREMISLKEI
1314 AAPVSFPTSTSSDSSGGRVIPDVRVTOHFAKEPQEPKLLKSHSPASP--SLASKEY
1092 CTGSNAT-----YHGAKGEHTSRKDAWTAQNTGISTLYRNSYGAPAI
1373 PGGTPATTAMAPCPASLPRGYMTMPAGERSPSTSTIHSYGQPTTAN--YGSQTH
1138 -----NQVSAQPV-----POEPSRKDYETY-----QPFQNSTRYNDESGFI
1432 PSGPAGSGRASREKFLGGDGEVGPQPSR--GYSYFTGSSPPLSPSPSPSPSPFIF
1178 HRPPASEYTHVLGLKSTGNVDFYSAARPYSELNYETSHYPASP 1221

```

RESULTS
 T29265
 hypothetical protein C01G8.7 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T29265
 R:Du, Z.; Gattung, S.
 submitted to the EMBL Data Library, November 1996
 A:Description: The sequence of C. elegans cosmid C01G8.
 A:Reference number: Z20597
 A:Accession: T29265
 A:Status: preliminary; translated from GB/EMBL/DBDJB
 A:Molecule type: DNA
 A:Residues: 1-1357 <DUZ>
 A:Cross-references: EMBL:U00439; PIDN:AAB37645.1; GSPDB:GN00019; CESP:
 A:Experimental source: strain Bristol N2; clone C01G8
 C:Genetics:
 A:Gene: CESP:C01G8.7
 A:Map position: 1
 A:Introns: 99/1; 488/1; 701/3; 1056/2; 1159/3; 1197/1; 1312/3
 Query Match 3.4%; Score 217; DB 2; Length 1357;
 Best Local Similarity 18.0%; Pred. NO. 0.0019;
 Matches 267; Conservative 162; Mismatches 516; Indels 540;

QY	6	PPGAAPL	-----	GAMPV	PDQSSASEKTSSLS	PG	-----	LNTE
Db	96	PPGGPQPF	PGHGMDF	PNHYYY	QQHGMMP	PHFGHP	GYPPQH	NMFSPGQYPGHQRP
QY	43	SETETTSAIL	ASVKEQEL	QFERLT	TRELEAER	QIVASQLER	CKLGSET	SGSMSSMS
Db	156	GPPPGQAR	PMQHMDE	HOQWHA	QAAAAH	----- <td>----- <td>HHZ</td> </td>	----- <td>HHZ</td>	HHZ
QY	103	QWQSDGQK	IDEELT	TGELVDS	CIRLSQ	SGLDPDY	STGTRPS	LLSQALQI
Db	200	AAQAQAAQ	----- <td>AAQAST</td> <td>PAP</td> <td>----- <td>STTPA</td> <td>QAASSQA</td> </td>	AAQAST	PAP	----- <td>STTPA</td> <td>QAASSQA</td>	STTPA	QAASSQA
QY	161	-----	KPGESFQ	-----	YPASYHS	NOTIL	----- <td>ALGETT</td>	ALGETT
Db	237	NTQATP	SASNP	PTPASAS	ASATL	THVEPAS	VAGSQAG	SRAPSGPSPATPDS
QY	193	TOARATG	SFSGGTT	SRAHLAG	PEPA	----- <td>PPP</td> <td>-----</td>	PPP	-----
Db	297	ASAAADA	QTTSST	STGPTST	SDTAP	STSTVP	PTTSTVP	QPPQPPQ
QY	240	LPDAPPA	AAAAAL	YSSSTL	PAPRGG	SPLAAP	QGSPTKLQ	-----
Db	347	-----	PGQPPG	----- <td>PGQHPG</td> <td>YPGYG</td> <td>YPGPG</td> <td>AMRPPAGPAPPGAPY</td>	PGQHPG	YPGYG	YPGPG	AMRPPAGPAPPGAPY
QY	297	GSSPKQ	PSRLAK	SYSTSS	PNIVV	SAGLSPI	RVTSP	PTVOSTISSPHQLSSI
Db	394	GAPP	----- <td>----- <td>PAGF</td> <td>HPHQHQAQ</td> <td>----- <td>-----</td> </td></td>	----- <td>PAGF</td> <td>HPHQHQAQ</td> <td>----- <td>-----</td> </td>	PAGF	HPHQHQAQ	----- <td>-----</td>	-----
QY	357	ATLSPT	KLVLHAS	BOYSKH	SQELVAT	TATLQ	PGSLAAG	SRASSYSQHQHILGPELRA
Db	414	-----	YLAWQ	QOQRYHQ	QOQOQOQ	----- <td>GAPG</td> <td>-----</td>	GAPG	-----
QY	417	EBHIDP	IEDRVY	QKPPMR	SLSQSG	DPLP	PAHTGT	-----
Db	457	QNRMP	----- <td>PP</td> <td>----- <td>PAQAPS</td> <td>PSGAAG</td> <td>SKGPKRYCTPAPPASASPT</td> </td>	PP	----- <td>PAQAPS</td> <td>PSGAAG</td> <td>SKGPKRYCTPAPPASASPT</td>	PAQAPS	PSGAAG	SKGPKRYCTPAPPASASPT
QY	470	QRTSGH	GPQNA	AATFOR	ASYAG	PASNVAD	PRQLOY	CPSVESYKSGKGPALPP
Db	503	SSTMEV	VAPST	STQPTT	TGTYL	----- <td>ANTLATP</td> <td>-----</td>	ANTLATP	-----
QY	530	ARSP	SIDS	IQKDP	REFGWR	DELPEV	IQMLQ	HPSPVQSNAALVLOHL
Db	543	SHA	----- <td>----- <td>HHVIS</td> <td>QOQHQP</td> <td>PPGCI</td> <td>EATATSQ</td> </td>	----- <td>HHVIS</td> <td>QOQHQP</td> <td>PPGCI</td> <td>EATATSQ</td>	HHVIS	QOQHQP	PPGCI	EATATSQ
QY	590	RQGGI	QILLV	DLDR	MT	EVHRS	----- <td>ACGAL</td>	ACGAL

ne PGM-2A) - pig (fragment)
 ofa domestica (domestic pig)
 7 #sequence_revision 21-Feb-1997 #text_change 03-Nov-2000
 1: S55315
 askar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
 06, 200, 1994
 ic mucin: isolation and characterization of a cDNA clone with a novel
 1: 147141; MUID:94102478; PMID:7506218
 RNA
 arY; translated from GB/EMBL/DBDJB
 <TUB>
 : EMBL:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208
 askar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
 9-96, 1995
 and characterization of cDNA clones encoding pig gastric mucin.
 : S55315; MUID:95275264; PMID:7755593
 5
 arY
 RNA
 <TUB>
 : GB:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208
 submaxillary mucin
 3.4%; Score 217.5; DB 2; Length 528;
 arity 23.9%; Pred. No. 0.00048;
 onservative 73; Mismatches 255; Indels 95; Gaps 22;
 PD-----QPSSA-----SEKTS-----LSPGLTNGSGSETETTSAILASV 55
 PSTTSVQPSSSGAPTTTSVOTSSSSPPISSTISVOTSSSSVPTTST--TSV 84
 LQFERLTRELEAERQIVASQLERCKLGSETSGSMSSAEEQFQMSQDQKDIED 115
 SSSAPITRATSVQ-----SSSSSAPISSTTSVQPSSSGVPITTSVQSSSSSA 140
 3ELVDSCTRISQESGILDPOYSTGERPSLLSQALQNSKPGSFQYPASYHNS 175
 ATSVQPSSSSSPPISSTTSVQPSSSSAPTTTSV-----QFSSSSGPPIS--- 191
 LGFTTSPQLPARQTQARATGQSFSGTTSRAHLAGPEPAPPPPPPPPPPPAPSLG 235
 VQTSSSSSVPTTSTTSVQPSSSSSVPTTTSATS--VRSSSSSTPTTSTTSVQPS 249
 LPDAPPAALYSSSTLPAPRGGSPPLAAPQGSPTKLQ--RGGGAP--EGATY 292
 ---APTTSATSVQPSSSSTPIP-----SITSVQPSSSSAPTTTSATS 291
 3SS--PKQPSRLAKSYSTSP-----INTVSSAGLSPIRVTSPPTVQSTISSSP 346
 SSSSPPISSITSVQPSSSSPTTSTTSVQPSSSGAP--TTSATSVQPSSSSSP- 348
 STIG---TYATLSPTKLVLHASEQYKSHQELVATLQRPQSLAAGSRASSVSH 403
 STISVQPSSSSSPTTST--TSVQPSSSGAPTTTSVQPSSSSVPTTTSATS-- 403
 PELRALQSPHEHIDPIEDRVYQKPPMRSLSQSGDELPEAHTGYRTSTAPSSPG 463
 -----VRSSSSSTPI-----PTTTSVQPSSSSVPTTTSATSVQTSSSSTPI 446
 PLQRTSGHGPQNAAATFORASYAGPASNVADPRQLOYCPSVESYKSGKGPAL 523
 SVQPSSSSSAP--TTSATSVQPSSSSSPPISSTI---SVQPSSSSSSTTSTTSVQ 501
 TLARSPIDSIO 539
 GAPTTSATSVQ 517


```

STFASSTPIASSSPGSTVTVAPG--SSSTYGSSTPSASSSSSGTMS-----T 412
AERQIVASQLERCKLGGSETGSMSSMAEEOQFOWQSDGQKDIEDELTTLGLVDS 126
TGSTVTVAPVSSSTFGSSTPIASSSSG-----S 445
LOESGILDPQDYTGTERPRLISQALQLNKSPGSPQYPASVHGNQTLALGETTFS 186
V--SGSSSTYGSST--PSASSSA-GTASTISGSTGSTATIVPGSSSSSVGSSTQS 499
R-GTQARATGQSFQSGTTSRAGHLAGEPAPPPPPPPPPPPPP-----232
SPGTMSTVSGTGTSTTV-----VPGSSTSPASSSPNPSSPASTGTTIIGSS 554
-----SLGSAFHLDPAPAAAAAALYSSSTLPAPPRGG-----266
STVSGSTVSGTGTSTQSLASSTATPSSSTVPSSSPQSSQSPAPNTGSTTSPSQ 614
--SPLAAPQGSPTKLQGGSAPEGATVAAPRGSS-----PKOSP 304
SPSPSMNPSSSTPTGSSQSTIIPRGSTASSPTGSTGTSTFVATEVTSQSTVPSSGS 674
KSYSTSPINIVSSAGLSPIRVTSPTVOSTISS-----SPIHQLSS--351
STNSSPSPSLSPTSGMSTL--TSEPSPSTQSSGAQSTLTTPSPNPSTSSLE 732
-----TIGTYATLSPTKRLVHASEQYS-----373
3ATTSSGAGTTWT--SPSQSSSVGSSGSGTSPAATSTSGEMTSQGSTQTPGSSVST 791
-KHSQELIAT----ATLQRP-----GSLAAGSRAYSYQHGHGELRALQSP---416
JTSTQOSVNSTSPGSTVTRPSTVSGTSSGSTVTVGSTASTSGSSVASSSPAPST 851
CPYI--EDRVYQKP--PMRSISQSGDPLP-PAHTGTYRTSTAPS-----460
VPTSSGSSMITQSPVPSQSTSPVESSTTPSPGSPGTTLTSTSPSPSQSTTIGSTQ 911
PGVDSVPLQRT--GSGHGPQNAATAAATPQRA---SVAAAGPASNVADPYRLQVCPS 511
PGISTTSEMTSQGSTQTPGSTVTVQSTVSDSTSSGSTVTVGSTGSSSIPPS 971
-SPYSKGPAL---PREGTLARPSIDSIQDPREFGWRDPPELPEVIMQLQHOFPSS 565
[NPSTSGSSMSTQTPQSSQSTSPVESSTSGATSSGSGSPGTTLTSTI-----SPSPS 1026
AAAYLQHLCPGDNKIAEIRROGGIQLLVLLDHRMTEVHRSACGALR--NLVYCK 623
[GS-----SQGSTSPVVSSTISQGSTETPGSTGTVTKPSTVSGS 1069
YKI-----ALKNCGGIPALVRLLRKTTDLLEIRELVTGLVNLSSCDALKMPIIQD 677
TATMGSTEAASSTSGSSSTSPNSQSTSP-----STSGATSSP-----1111
JTNVILPHSGWENSPLODDRKLQLHSSQVLRNATGCLRNV---SSAGEARRMRRE 735
TLTSTISP-----SPSQ-----SSTIGSSQGSTSPVVSSTTSGDMTSQSTQ 1156
DALLYVIOALGSSBIDSK---TVENCVCILR--NLSYRLAAETSGQHMGTDELD 791
GST--VTQPSGTSGSTSGEITSQSTQTPRESSLSTSPALSTSTQSSVSTNSPG 1214
EANGKDAESSGCW-----GKXKKKKSQDQWDGVLPLPCAEPPKGI-----838
PSTVARGSTSGSTVTTGTEGSSSTSGSSSATLSLSSSPVPSSTSQSPNSTSGSST 1274
-----QMLWPSIVKPYLTLLSECSNPDTEGAAGALQNLAAQ-----876
SOSTSPVVSSTTGTGEMTSHGSTQTP-STIGSTVTPQSTVSGSSSGSTVTIGSSEA 1333
-SWKSVYIRAAVRKEKGLPILVELLRIDNDRVVCVATALRNALDVRNKELIGK 931
ISFKTSPSSISPVPTSSPIP-----STTFASSTSGSTISDVSSVST 1378

```

```

QY 932 YAMRDVLHRLPGG-----NNSNNTASKAMSDDTVTAVCTLHEVITKMNENAI
Db 1379 TSLAPLSSSLSTVPSTQSFSTSGSSKASSSPVPSQSTSTPTNP-TGSTESS
QY 984 A-GGIEKLVGISKSGDKHSPKVVKAASQVLNSMWQYRDLRLSLYKKGQWSQVHFVI
Db 1438 TISGSTQHTTMSKASSGSTSPSTNSQSTVTM-----GSSSTSGVE
QY 1043 IERDRQPPYS--SRTPSISPVRVSPNNRSASAPASPREMIS-----LKERP
Db 1484 SSTQPMSTSGSSAGSTVASSTASP-AASSTAPSSSTGTMSTSGSTVGSTISESE
QY 1092 C---TGSNATVHGAKGHEHTSPKDAWTAQNTGISTLYRNSYCAPAEDIKHNVOVSQF
Db 1543 ASSQTGSTVTWSSSTSGVSTSSASTQ-PQMSTSGSSAGSTVASSTAGIVSTSI

```

RESULT 14
T49726
hypothetical protein B23L21.390 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49726
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holla
submitted to the Protein Sequence Database, May 2000
A:Reference number: 225022
A:Accession: T49726
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1198 <SCH>
A:Cross-references: EMBL:AL356172; GSPDB:GN00116; NCSP:B23L21.390
A:Experimental source: BAC clone B23L21; strain OR74A
C:Genetics:
A:Gene: NCSP:B23L21.390
A:Map position: 6
A:Introns: 10/3; 188/3

Query Match 3.3%; Score 207.5; DB 2; Length 1198;
Best Local Similarity 22.6%; Pred. No. 0.005;
Matches 141; Conservative 69; Mismatches 245; Indels 169; G

```

QY 7 PGAAPLGAM---PVPDPQSSASEKTSLSPLNLS-NGDGETETTTSA---ILAS
Db 367 PGGVELGSIEKRPMP--PSRAGTFATQFSGASSTSLIGAAAEAPARQANDAPTVPK
QY 59 ELQFERLTRELEAERQIVASQ--LERCKLGSSETGSMSSMAEEOQFOWQSDGQKD
Db 425 NIPPTRTATMLSQSSYGTASQSNFEPRLPGLAAATGLSNYDAASSQSNFQPRPIR-
QY 117 LTTGLEAVDSICBSIQESGILDQDYSTGERPSLLSQSALQLNSKPEGSFYQYASY
Db 480 TATGISHVSS-VPSQQPS-----FQPRPLTRAATGMSNMGEGP---SQSF
QY 177 TLALGE--TTPSQLPARGTQARATQSFQSGTTSRAGHLAGEPAPPPP-----PPPR
Db 525 PRELTRAATGLSNTSAGPQSQADFEPPRACTVPPQNNFRFCMQQPPPPPRAGTAPPR
QY 232 PSL---GSAFHLDPAPAAAAAALYSSSTLPA---PPRGSPPLAAPQGSSTKLQ
Db 585 PSMQPNSTF-----GPGYTDSPATYSTATMPAFQLPAR--APTAPPMYSNNNY-
QY 286 APEGATVAAPRGSSPKQSPSRKLSVSTSSPINIVSSAGLSPIRVTSPTVOSTI:
Db 633 -----RGVEPR--PIERAYTGNVGNQSHDNFSAG-----
QY 346 IHQLSSTIGTYATLSPTKRLVHASEQYSK-----HSQELYATATLQRPGLAAGSI
Db 661 -NQVPTNPNYGGFSG-----HMSNCYDDGGYGGYQDIYQDNYGQQGYMNGM
QY 400 SSOQHGLG-----PELRLQSPHEHI-----DPIYEDI
Db 712 HDNYGHDGYPQANNQSPRLGSAAPPQTQSGYQPYIPQRTMTAPPBQPYHHK

```

14:26:20 2004

us-09-501-171a-4.rpr

```

1076 SPREMISLKERKTDY-----ECTGS-----NATYHGAKGEHTSRKDMATAQNTGI
113 AVLFKMS-EDKSQDYKKRLSVELTSSLFRDDVPWEPG---NTMADILTSQ-----
1126 NSYGAPAEIDIKHNOVSAQVPOEPSRKDYETQPPQFNSTRNVDSEFFE 1173
1760 -SY---ADBIYSPHVSQNNLSYNPNYSYHQHSGMFPQMONNVTOGWFD 803

Search completed: April 22, 2004, 12:15:23
Job time : 34 secs

MRSLS-----SQGDPLPPAHTGT-----YRTSTAPSPGVDSVPLQRT 472
KSSLSVAELPAELPVEISSEQLPAPAGSVTSYLEYMDGSPSSRRITDAPPKTRT 829
HGPNAAATFO-----RASYAAGPASNYADPYRQLQVCPSPVESPYSK 518
SSQGSAAATQETKSSGSHRQSGAGSVSKAAGAGSNSTSSRK-----GSVASSVSS 885
ALPEGTLARSPSIDSTOKOP 542
SHQKQSGSVASVSTSSRRDP 909

dra magnipapillata
magnipapillata
96 #sequence_revision 15-Oct-1996 #text_change 21-Jul-2000
35
atta, M.; Fischer, R.; Fujisawa, T.; Holstein, T.W.; Sugiyama, T.
cation of a Hydra homologue of the beta-catenin/plakoglobin/armadillo
r: JC4835; MUID:96257271; PMID:8654977
35
acid sequence not shown
nrna
<HOB>
s: GB:U36781; NID:G1407600; PIDN:AAC47137.1; PID:G1407601
rotein plays roles in cadherin-mediated cell adhesion and in signal tra

3.2%; Score 207; DB 2; Length 806;
larity 21.5%; Pred.No.0.0031;
Conservative 110; Mismatches 233; Indels 166; Gaps 29;
VIOMLQHOFPSVQSNAAAYLOHLCFNGNKIKAEIRROGGIQLLVLDLHDMTEVHRS 611
LVKLLGHRIEAVVFYAITTLHNLHLLHQEGAKMAVRLALGLQKMYSLIQRPKVKFLAI 355
ALRNLVYGKANDDKIALKNCGGIPALVRLIRKTTDLRELYTGVLNLSGCDALK 671
CLQLIAYG--NQESKLILSSGGPAELVRIMRSYTEKLLYTTCRVLVKLVSVCSNNK 413
IQDALAVLTNVAIIPHSGWENSPLODDRKIQLHSSQVLRNATGCLNRVSSAGEARR 731
IVEAGGQALAHYLSH-----QSTRLVQNCWLTNRNLSDVATKQ-- 454
ECDGLTDALLYVTQSALGSSEIDSKTVENCVC-ILRNLRYRLAAETSQGOHMGTDDEL 790
--DGL-EGLLQMLVQLLSSNDIN--VVTCSVGIISNLT----- 487
LCGEANGKDAESSGCGWKKKKKQDQWGVGLPDPDCAEPPKGIQMLWHPISVKPY 850
-----CNPKNKQVVFQ-----VG-----GIEAL-----V 507
LSECSNPDTLEGAGALQNLAAAGSKWVYIRAAVRKEKGLPILVELLRIDND-RVY 909
INAGDREITEPAVCALRHLS-RHPDAEHAENGVRHLYGIPILVKLLNPPSRWPLI 566
ATALRNMAL-----DVRNK-----ELIGKYAMRDLVHRLPGNN-SNNTASKAM 955
VGLIRNLGLCFSNHTPIRDQGLPKLVQLLMK-SYQDIQRGFGAQMDOGVRMEEI 625
IVTAVCCTLHEVITKMNNAKALRDAGGIEKLVGISKSGDKHSPKVVKAAASOVLS 1015
TVGA-----LH-ILAREALNRSIIDLNCITFFVQLLYSEV-----NIVRAAGVLCE 676
YRDLRSLYKDGWSQYHFVASSSTIEDRQRPVSSSTPTSPVSPVSPNNRSASAPA 1075
-----DKEG-----ADAIERE-----GATTITELLSRNDGIAAYAR 712
```

```

1076 SPREMISLKERKTDY-----ECTGS-----NATYHGAKGEHTSRKDMATAQNTGI
113 AVLFKMS-EDKSQDYKKRLSVELTSSLFRDDVPWEPG---NTMADILTSQ-----
1126 NSYGAPAEIDIKHNOVSAQVPOEPSRKDYETQPPQFNSTRNVDSEFFE 1173
1760 -SY---ADBIYSPHVSQNNLSYNPNYSYHQHSGMFPQMONNVTOGWFD 803

Search completed: April 22, 2004, 12:15:23
Job time : 34 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

in search, using sw model
ril 22, 2004, 12:06:54 ; Search time 19 Seconds
(without alignments)
3357.155 Million cell updates/sec

-09-501-171A-4
77
MEARKPPGAAPLGAAMPVQD.....PYSELNYETSHYPASPDWV 1225

OSUM62
pop 10.0 , Gapext 0.5

1681 seqs, 52070155 residues

ts satisfying chosen parameters: 141681

gth: 0
gth: 2000000000

inimum Match 0%
aximum Match 100%
isting first 45 summaries

wiseProt_42:*

the number of results predicted by chance to have a
r than or equal to the score of the result being printed,
ed by analysis of the total score distribution.

SUMMARIES

ery	ch	Length	DB	ID	Description
9.5	1225	1	CTD2_HUMAN	Q9ubq3	homo sapien
6.1	1247	1	CTD2_MOUSE	Q35927	mus musculus
2.3	1211	1	PKP4_HUMAN	Q99569	homo sapien
2.3	969	1	ARVC_MOUSE	P98203	mus musculus
2.2	962	1	ARVC_HUMAN	O00192	homo sapien
1.6	911	1	CTD1_MOUSE	P30999	mus musculus
1.5	264	1	CTD2_RAT	Q35116	rattus norv
1.5	968	1	CTD1_HUMAN	O80716	homo sapien
2.6	881	1	PKP2_HUMAN	Q99959	homo sapien
2.1	797	1	PKP3_HUMAN	Q9y446	homo sapien
2.0	797	1	PKP3_MOUSE	Q9gy23	mus musculus
1.1	728	1	PKP1_MOUSE	P97350	mus musculus
1.0	747	1	PKP1_HUMAN	Q13835	homo sapien
0.8	727	1	PKP1_BOVIN	Q28161	bos taurus
4.0	1185	1	DRPL_HUMAN	P54259	homo sapien
3.6	1183	1	DRPL_RAT	P54258	rattus norv
3.5	1324	1	SAL1_HUMAN	Q9nsc2	homo sapien
3.5	802	1	ENAH_MOUSE	Q03173	mus musculus
3.5	1300	1	SAL3_HUMAN	Q9bxa9	homo sapien
3.2	721	1	YK82_MYCTU	Q10690	mycobacteri
3.2	5560	1	SPEN_DROME	Q8x883	drosophila
3.2	1322	1	SAL1_MOUSE	Q9er74	mus musculus
3.2	5038	1	PCLO_MOUSE	Q9gyx7	mus musculus
3.2	1487	1	ICP4_HSVB	P28925	equine heip
3.2	817	1	VRP1_YEAST	P37370	saccharomyc
3.1	1902	1	SNF1_HUMAN	O4497	homo sapien
3.1	2845	1	APC_MOUSE	Q61315	mus musculus
3.1	1461	1	IE18_PRVIF	P11675	pseudorabie
3.1	741	1	GTSE_MOUSE	Q8r080	mus musculus
3.1	1140	1	YM96_YEAST	O04893	saccharomyc
3.1	2716	1	OSA_DROME	Q8in94	drosophila
3.1	2842	1	APC_RAT	P70478	rattus norv
3.0	660	1	YHL1_EBV	P03181	epstein-bar

34	194	3.0	1575	1	SVJ1_HUMAN	O43426	homc
35	194	3.0	2843	1	APC_HUMAN	P25054	homc
36	193.5	3.0	781	1	CTNB_RAT	Q9wu82	rati
37	193	3.0	1323	1	SAL3_MOUSE	Q62255	mus
38	192.5	3.0	4911	1	MLL3_HUMAN	Q8ne24	homc
39	191.5	3.0	781	1	CTNB_HUMAN	P35222	homc
40	191.5	3.0	1083	1	T2D3_HUMAN	O00268	homc
41	190.5	3.0	1744	1	TENS_CHICK	Q04205	gal.
42	190	3.0	1087	1	ASPI_MOUSE	Q62415	mus
43	190	3.0	1487	1	ICP4_HSVK	P17473	equi
44	189.5	3.0	781	1	CTNB_MOUSE	Q02248	mus
45	188.5	3.0	1273	1	WEB1_YEAST	P38968	sacc

ALIGNMENTS

RESULT 1
CTD2_HUMAN
ID_CTD2_HUMAN STANDARD; PRT: 1225 AA
AC Q9UQ83; C00379; O15390; O43206; O43840; Q13589; Q9UM66; Q9UPM3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Catenin delta-2 (Delta-catenin) (Neural plakophilin-related ARM-i
protein) (NPRAP) (Neurojuncin) (CT24).
GN CTNND2 OR NPRAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH PRESENILIN 1
MEDLINE=97366296; PubMed=9223106;
RA Zhou J., Liyanage U., Medina M., Ho C., Simmons A.D., Lovett M.,
Kosik K.S.;
RT "Presenilin 1 interaction in the brain with a novel member of the
RT Armadillo family."
RL NeuroReport 8:2085-2090(1997).
[2]
SEQUENCE FROM N.A. (ISOFORM 1), AND FUNCTION.
TISSUE=Brain;
MEDLINE=99268159; PubMed=9971746;
RA Lu Q., Paredes M., Medina M., Zhou J., Cavallo R., Peifer M.,
Orecchio L., Kosik K.S.;
RT "Delta-catenin, an adhesive junction-associated protein which pre
cell scattering."
RL J. Cell Biol. 144:519-532(1999).
[3]
SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH PRESENILIN 1
TISSUE=Fetal brain;
MEDLINE=99223289; PubMed=10208590;
RA Tanahashi H., Tabira T.;
RT "Isolation of human delta-catenin and its binding specificity wit
presenilin 1."
RL NeuroReport 10:563-568(1999).
[4]
SEQUENCE OF 186-1225 FROM N.A. (ISOFORM 1), AND INTERACTION WITH
PRESENILIN 1.
TISSUE=Brain;
MEDLINE=99155075; PubMed=10037471;
RA Levesque G., Yu G., Nishimura M., Zhang D.M., Levesque L., Yu H.,
Xu D., Liang Y., Rogava E., Ikeda M., Dutchie M., Margolo N., Wan
VanderVere P., Bayne M.L., Strader C.D., Rommens J.M., Fraser P.E
St George-Hyslop P.H.;
RT "Presenilins interact with armadillo proteins including neural-
specific plakophilin-related protein and beta-catenin."
RL J. Neurochem. 72:999-1008(1999).
[5]
SEQUENCE OF 350-1225 FROM N.A. (ISOFORM 1), AND ALTERNATIVE SPLIC
TISSUE=Fetal brain;
MEDLINE=98002299; PubMed=9342840;
RA Paffenholz R., Franke W.W.;

ion and localization of a neurally expressed member of the
"armadillo multigene family";
ion 61:293-304(1997).

543-656 FROM N.A.
Smith A., Elliott G., Kramer J., Latreille P., Keppler D.;
JOV-1997) to the EMBL/GenBank/DBJ databases.

865-1225 FROM N.A.
2103; PubMed=9049630;
Overhauser J., Lovett M.;
of cDNAs from the Cri-du-chat critical region by direct
a chromosome 5-specific cDNA library";
7:118-127(1997).

873-1225 FROM N.A.

son J., Gibbs R.A.;
JOV-1997) to the EMBL/GenBank/DBJ databases.
l: May be involved in neuronal cell adhesion and tissue
mesis and integrity by regulating adhesion molecules.
cell motility and cell scattering in response to
re growth factor treatment (in vitro).

BINDS TO E-CADHERIN AT A JUNCTIONAL SITE WITHIN THE
MIC DOMAIN (BY SIMILARITY). BINDS TO PRENILIN-1.

ILAR LOCATION: Adherens Junction (By similarity).

IVE PRODUCTS:

ternative splicing; Named isoforms=2;

t=Additional isoforms seem to exist;

Synonyms=NPRAPA;

Q9UBQ3-1; Sequence=Displayed;

Synonyms=NPRAPC;

Q9UBQ3-2; Sequence=VSP_006746;

PECIFICITY: Predominantly expressed in brain.

TY: Contains 9 ARM repeats.

TY: Belongs to the beta-catenin family.

REF 8 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS

TON 746 TO 913.

ROT entry is copyright. It is produced through a collaboration
Swiss Institute of Bioinformatics and the EMBL outstation -
Bioinformatics Institute. There are no restrictions on its
-profit institutions as long as its content is in no way
this statement is not removed. Usage by and for commercial
quires a license agreement (See <http://www.isb-sib.ch/announce/>
mail to license@isb-sib.ch).

; AAC63103.1; -

; AAB96357.1; -

05; BAA36163.1; -

; AAD00453.1; -

; AAB68599.1; -

; AAB97957.1; -

89; AAB33940.1; -

02; AAB88185.1; ALT_FRAME.

2516; CTNND2.

58; P:neuronal cell adhesion; NAS.

R008938; ARM.

R000225; Armadillo.

4; Armadillo_seg; 5.

85; ARM; 7.

0176; ARM_REPEAT; 3.

n; Cytoskeleton; structural protein;

1 protein; Repeat; Coiled coil; Alternative splicing.

49 84 COILED COIL (POTENTIAL).

216 226 POLY-PRO.

811 817 POLY-LYS.

394 438 ARM 1.

540 579 ARM 2.

582 621 ARM 3.

626 666 ARM 4.

682 724 ARM 5.

FT	REPEAT	728	773	ARM 6.
FT	REPEAT	835	875	ARM 7.
FT	REPEAT	882	921	ARM 8.
FT	REPEAT	975	1018	ARM 9.
FT	VARSPLIC	823	880	Missing (in isoform 2).
FT	CONFLICT	426	426	/FTid=VSP_006746.
FT	CONFLICT	443	443	V -> D (IN REF. 3, 4 AND 5).
FT	CONFLICT	725	725	P -> A (IN REF. 3, 4 AND 5).
FT	CONFLICT	894	894	R -> L (IN REF. 3, 4 AND 5).
FT	CONFLICT	909	909	A -> V (IN REF. 3, 4 AND 5).
FT	CONFLICT	980	980	A -> G (IN REF. 5).
FT	CONFLICT	1102	1102	A -> G (IN REF. 4).
SEQ	SEQUENCE	1225	1225	AA; 132664 MW; 8B676CDFS4FA4E65 CRC64;

Best Local Similarity	Score	DB 1;	Length
99.58;	99.64;	DB 1;	Length 1225;
Matches 1220;	Conservative	0;	Mismatches 5; Indels 0;

Qy	1	MFARKPGAAPLGAMPVPDQPSASEKTSLSPLGNTSNGDSETEITTSAILAS
Db	1 <td>MFARKPGAAPLGAMPVPDQPSASEKTSLSPLGNTSNGDSETEITTSAILAS</td>	MFARKPGAAPLGAMPVPDQPSASEKTSLSPLGNTSNGDSETEITTSAILAS
Qy	61	QFERLTRELEAERQIVASQERKLGSETGSMSSMSAEQFQWQSDGQKDIEI
Db	61	QFERLTRELEAERQIVASQERKLGSETGSMSSMSAEQFQWQSDGQKDIEI
Qy	121	LELVDSICIRSLQSGILDPODYSTGERPSLLSOSALQNSKPEGSFOYBASYHS
Db	121	LELVDSICIRSLQSGILDPODYSTGERPSLLSOSALQNSKPEGSFOYBASYHS
Qy	181	GETTSPQLPARGTQARATGQSFQGTTSRAGHLAGPEPAPPPPPPPPPPPPPS
Db	181	GETTSPQLPARGTQARATGQSFQGTTSRAGHLAGPEPAPPPPPPPPPPPPS
Qy	241	PDAPPPAAAAAALYSSSTLPAPPRGSGPLAAPGQSGPTKLQSGSAPEGATYAAI
Db	241	PDAPPPAAAAAALYSSSTLPAPPRGSGPLAAPGQSGPTKLQSGSAPEGATYAAI
Qy	301	KQSPRLAKSYSTSSPINIVSSAGLSPIRVTSPTVQSTISSPHQLSSTIGT
Db	301	KQSPRLAKSYSTSSPINIVSSAGLSPIRVTSPTVQSTISSPHQLSSTIGT
Qy	361	PTKRLVHASQYKSHSQELVATATLQRPGLAAGSRASYSQHGHLGPELRLAQ
Db	361	PTKRLVHASQYKSHSQELVATATLQRPGLAAGSRASYSQHGHLGPELRLAQ
Qy	421	DPYEDRVYQKPMRSLQSQGDPLPPAHTGTTRTSTAPSSPGVDSVPLQRTGSC
Db	421	DPYEDRVYQKPMRSLQSQGDPLPPAHTGTTRTSTAPSSPGVDSVPLQRTGSC
Qy	481	AAAAATFORASYAAGPASNYADPYRQLQYCPSPVSPYKSGPALPPEGLTARSFI
Db	481	AAAAATFORASYAAGPASNYADPYRQLQYCPSPVSPYKSGPALPPEGLTARSFI
Qy	541	DPREFGWRDPPELPEVIQMLQHPSPVQSNAAAYLQHLQHLQHLQHLQHLQHLQ
Db	541	DPREFGWRDPPELPEVIQMLQHPSPVQSNAAAYLQHLQHLQHLQHLQHLQHLQ
Qy	601	LDHRMTEVHRSACGALRNLYGKANDDNKIALKNCGGIPALVLLRKTDTLEIRE
Db	601	LDHRMTEVHRSACGALRNLYGKANDDNKIALKNCGGIPALVLLRKTDTLEIRE
Qy	661	LWNLSGCDALKNPFIQDALAVLTNAVII PHSGWENSPLODDRKIQHSSQVLRNA
Db	661	LWNLSGCDALKNPFIQDALAVLTNAVII PHSGWENSPLODDRKIQHSSQVLRNA
Qy	721	NVSSAGEEARRRMRECDGLTDALLYVIQSLGSSSEIDSKTVENCVCILRNLSYRL
Db	721	NVSSAGEEARRRMRECDGLTDALLYVIQSLGSSSEIDSKTVENCVCILRNLSYRL
Qy	781	QGOHMTDDELGLLCEANGKDAESSGCGKKKKKKKKKKKKKKKKKKKKKKKKKKKK
Db	781	QGOHMTDDELGLLCEANGKDAESSGCGKKKKKKKKKKKKKKKKKKKKKKKKKKKK

MCMTDELGLL	CCEANGKDAESSGCGKXKKKKXQDWDGVGLPDCAEPPGIQM	840
STIVKPYLTL	LLSCSNPDTLEGAAGALQNLAAAGSKWKSIVYIRAAVRKEKGLPILVEL	900
STIVKPYLTL	LLSCSNPDTLEGAAGALQNLAAAGSKWKSIVYIRAAVRKEKGRPIILVEL	900
NDNRVVCATAT	RNALDVRNKELIGKYAMEDLVHRLPGGNNNNNTASKAMSDDTV	960
NDNRVVCATAT	RNALDVRNKELIGKYAMRDLVHRLPGGNNNNNTASKAMSDDTV	960
CTLHEVITKN	MENAKALRDAGGIEKLVGISKSGDKHSPKVKAAASQVLNSMWQYR	1020
CTLHEVITKN	MENAKALRDAGGIEKLVGISKSGDKHSPKVKAAASQVLNSMWQYR	1020
LYKKDGMQSY	HFVASSSTIERDRQRPYSSSRSTPISIPRVYSPNNRSASAPASPREM	1080
LYKKDGMQSY	HFVASSSTIERDRQRPYSSSRSTPISIPRVYSPNNRSASAPASPREM	1080
ERKRTDYE	CTGSGNATVHGAKGHTSRKDAQTGISTLYRNSYGAPAEIDIKHNOV	1140
ERKRTDYE	CTGSGNATVHGAKGHTSRKDAQTGISTLYRNSYGAPAEIDIKHNOV	1140
WQPSRKDYTY	QPFQNSTRYNDESFEDQVHHRPPASEYTMHLGLKSTGNYVDF	1200
WQPSRKDYTY	QPFQNSTRYNDESFEDQVHHRPPASEYTMHLGLKSTGNYVDF	1200
RPYSELNVET	SHYPASPDWSV	1225
RPYSELNVET	SHYPASPDWSV	1225
STANDARD;	PRT; 1247 AA.	
Rel. 41,	Created	
Rel. 41,	Last sequence update)	
Rel. 42,	Last annotation update)	
2	(Neural plakophilin-related ARM-repeat protein)	
	(juncin)	
	ND2 OR NPRAP.	
	(Mouse).	
	azoz; Chordata; Craniata; Vertebrata; Euteleostomi;	
	eria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
	990.	
	N.A. (ISOFORM 1), AND ALTERNATIVE SPLICING.	
	orain;	
	99; PubMed=9342840;	
	Frank W.W.;	
	on and localization of a neurally expressed member of the	
	madillo multigene family.";	
	on 61:293-304 (1997).	
	INTERACTION WITH E-CADHERIN.	
	158; PubMed=9971746;	
	s M., Medina M., Zhou J., Cavallo R., Peifer M.,	
	Kosik K.S.;	
	n, an adhesive junction-associated protein which promotes	
	g.;	
	144:519-532 (1999).	
	SPLICING (ISOFORMS 1 AND 2), AND INDUCTION.	
	270; PubMed=10626844;	
	Fan Q.W., Hayashi H., Michikawa M., Yanagisawa K.,	
	f the mRNA for two isoforms of neural plakophilin-related	
	tein/delta-catenin in rodent neurons and glial cells.";	
	c. 27:185-188 (1999).	
	ICITY, AND SUBCELLULAR LOCATION.	
	926; PubMed=10753311;	

```

247 AA; 134998 MW; D4YA7A6B6A27D2919 CRC64;
/FTId=VSP_006747.
96.1%; Score 6129; DB 1; Length 1247;
larity 94.8%; Pred. No. 2e-283;
Conservative 8; Mismatches 29; Indels 28; Gaps 3;

RKPGAAPLAMPVPDQPSASAKTSLSPLGINTSGDSETEITSAIASVKEQEL 60
RKQSGAAPFGAMPVPDQPSASEKNSLSPLGINTSGDSETEITSAIASVKEQEL 60
RLTRELEAERQIVASQLERCKLGETSGSMSSMAEEOFWQSGQKQIDELTTG 120
RLTRELEAERQIVASQLERCKLGETSGSMSSISAGEQFHWQIQGQKQIDELTTG 120
VDCIRSLOESGILDPQDYTCGRPSLLSQSALQNSKPEGSFOYPASYHSNQTAL 180
VDCIRSLOESGILDPQDYTS*SERPSLLSQSALQNSKPEGSFOYPASYHSNQTAL 180

TPSQLPARGTQARATGQSFSGTTSRAGHLAGPEPAPPPPPPPPPPPPSLGSFHL 240
APSQLPARSTQARAAGQSFSGTTRAGHLAGEPA-PPPPPPPPPPPPPSLGSFHL 239
PPAAAAALYYSSSTLPAPRGSGPLAAPQCGSPTKLQRGSAPEGATYAAPRGSSP 300
PP--AAAALYYSSSTLPAPRGSGPLTTQCGSPTKLQRGSAPEGATYAAPRGSSP 297
PSRLAKSYSTSPINIVSSAGLSPIRVTSPTVQSTISSPIHQLSSTIGTYATLS 360
PSRLAKSYSTSPINIVSSAGLSPIRVTSPTVQSTISSPIHQLSSTIGTYATLS 357
RLVHASQYSHSGLVATATLQPGSLAAGSRASYSSQHGHLGPELRALQSPHHI 420
RLVHASQYSHSGLVATATLQPGSLAAGSRASYSSQHGHLAPELRALQSPHHI 417
YEDRVYQKPPMRSLSQSGDPLPAHTGTYRTSTAPSGVDSVPLQRTGSHQPN 480
YEDRVYQKPPMRSLSQSGDPLPAHTGTFTSTAPSGVDSVPLQRTGSHQPN 477
ATFORASYAAGPASNADPYRQLYCPSPVESPYKSGPALPPEGLTARSIDSIOK 540
ATFORASYAAGPASNADPYRQLYCASVDSPYKSGPALPPEGLTARSIDSIOK 537
EFGWRDELPEVIQMLQHPFVSQSNAAAYLQHLQFQGNKIKABIRQGGIQLLVDL 600
EFGWRDELPEVIQMLQHPFVSQSNAAAYLQHLQFQGNKIKABIRQGGIQLLVDL 597
RMTEVHRSACGALRNLYGKANDDNKIALKNCGGIPALVRLRKTTOLEIRELVTGV 660
RMTEVHRSACGALRNLYGKANDDNKIALKNCGGIPALVRLRKTTOLEIRELVTGV 657
LSSCDALKMPIIQDALAVLTNAVITPHSGWENSPLODDRKIQLHSSQVLRNATGCLR 720
LSSCDALKMPIIQDALAVLTNAVITPHSGWENSPLODDRKIQLHSSQVLRNATGCLR 717
SAGEARRRRECECDGLTALLYVQISALGSEIDSKTVENCVCILNLSYRLAETS 780
SAGEARRRRECECDGLTALLYVQISALGSEIDSKTVENCVCILNLSYRLAETS 777
HMGDELDELGLLCGFRANGKDAESSCGWKKKKKKKQDQWGVGVLPCAPPPKGIOM 840
HMGDELDELGLLCGETNGKDTRESSCGWKKKKKKKQDQWGVGVLPCAPPPKGIOM 837
PSIVKPYLTLLSECSNPDTELEAGAGALQNLAAAGSWK----- 879
PSIVKPYLTLLSECSNPDTELEAGAGALQNLAAAGSWKGAEDVAGMAYALRSLPEGAP 897
-WSVYIRAAVKEKGLPLVELLRIDNDRVVCATATLRNMLDVRNKLIGKYAMR 935
QMSVYIRAAVKEKGLPLVELLRIDNDRVVCATATLRNMLDVRNKLIGKYAMR 957
HRLPGGNNNNNTASKAMSDDTVTAVCTTLEHVTIKNNMENAKALRDAGGIEKLVGISK 995

```

958 DLVHRLPGGNNNNSKAMSDDTVTAVCTTLEHVTIKNNMENAKALRDAGGIEK

996 SKGDKHSKPKVVVKAASOVLSNMWQYRDLRSLYKKQWSQYHFVASSSTIERDRQRI

1018 SKGDKHSKPKVVVKAASOVLSNMWQYRDLRSLYKKQWSQYHFVASSSTIERDRQRI

1056 TPISPVVSPVSPNNRSASAPSPREMISLKERKTDYECTSGSNATYHGAKEGHTSRP

1078 TPISPVVSPVSPNNRSASAPSPREMISLKERKTDYECTSGSNATYHGAKEGHTSRP

1116 QNTGISTLYXNSYGAEADIKHNOVSQAQVPOPSKDYETQYQPFQNSITRYDEK

1138 QNTGISTLYXNSYGAEADIKQNOVSTQVPOPSKDYETQYQPFQNSITRYDEK

1176 VHRPPASEYTMHLGLKSTGNYVDYFSAARPYSELNVTSHYPASPDQSVW 122E

1198 VHRPPASEYTMHLGLKSTGNYVDYFSAARPYSELNVTSHYPASPDQSVW 1247

RESULT 3

PKP4 HUMAN

ID PKP4 HUMAN STANDARD; PRT; 1211 AA.

AC Q99569;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Plakophilin 4 [p0071].

GN PKP4.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID:9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT), AND VARIANT SER-44

RC TISSUE=Frontal cortex.

RX MEDLINE=97092329; PubMed=8937994;

RA Hatzfeld M., Nacheheim C.;

RT "Cloning and characterization of a new armadillo family member,

RT associated with the junctional plaque: evidence for a subfamily

RT closely related proteins.";

RL J. Cell Sci. 109:2767-2778(1996).

CC -I- FUNCTION: May play a role in junctional plaques.

CC -I- SUBCELLULAR LOCATION: Colocalized with desmoplakin at desmos

CC junctional plaques in cultured epithelial cells.

CC -I- ALTERNATIVE PRODUCTS:

CC Event-Alternative splicing; Named isoforms=2;

CC Name=Long;

CC IsoId=Q99569-1; Sequence=Displayed;

CC Name=Short;

CC IsoId=Q99569-2; Sequence=VSP_006737;

CC -I- SIMILARITY: Belongs to the beta-catenin family.

CC -I- SIMILARITY: Contains 9 ARM repeats.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a col

CC between the Swiss Institute of Bioinformatics and the EMBL ou

CC the European Bioinformatics Institute. There are no restric

CC use by non-profit institutions as long as its content is i

CC modified and this statement is not removed. Usage by and for

CC entities requires a license agreement (See <http://www.isb-sib.ch>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; X81889; CAAS7478.1; -.

CC Genew; HGNC:9026; PKP4.

CC MIM; 604276; -.

CC GO; GO:0005911; C:intercellular junction; TAS.

CC InterPro; IPR008938; ARM.

CC InterPro; IPR000225; Armadillo.

CC Pfam; PF00514; Armadillo_seg; 4.

CC SMART; SM00185; ARM; 6.

CC PROSITE; PS0176; ARM_REPEAT; 3.

CC Cell adhesion; Cytoskeleton; Structural protein; Repeat; Coiled

CC Alternative splicing; Polymorphism.

CC -----

CC DOMAIN 36 70 COILED COIL (POTENTIAL).

115 455 ARM 1.
 116 557 ARM 2.
 117 599 ARM 3.
 118 644 ARM 4.
 119 702 ARM 5.
 120 751 ARM 6.
 121 855 ARM 7.
 122 901 ARM 8.
 123 993 ARM 9.
 124 794
 125 1085
 126 48
 127 448
 128 11 AA; 134268 MW; 14C626914A7024F CRC64;
 129 44.2%; Score 2818.5; DB 1; Length 1211;
 130 49.2%; Pred. No. 28-126;
 131 conservative 192; Mismatches 337; Indels 109; Gaps 37;
 132 DQPSASEKTSLSPLNTSGDSETETT-SAILASVKEQLOFERLTRELEAER 73
 133 EQASIVEBQQQTQEA-ASTGPGMEPETTATTILASVKEQLOFERLTRELEVER 59
 134 SOLERCKLSTGSMNSSSAEQFQMSOD-----GQKIDELTTGLELVDSC 127
 135 SOLERCLGABSPSTASTSTSEKFPNKRSTDVNTGVSKPRVSDAVQPNVYL---- 115
 136 QESGIL-DPQDYSYGERPSLL--SOSALOLNSKEGFSQFYPASVHSQNTLA----- 179
 137 PEQTLVSPETSUHESEGSIGNRSSTOMNSYSDSGYQEAGSFHNSQNVSKADNR 175
 138 -LGSTTPSQLPARGTOARATQCSFSQ-GTTSRAGHLAGPAPPPPP----- 224
 139 FIGSTNNHV-----RNSRAGQTLVQPSVANRVRVSSVPSRAQSSYVISTGV 231
 140 PFAPSLGSAFLP---DAPFAAAALYYSSSTLPAPPRGSPPLAAPQGGSPTKLQ 281
 141 SLRTSLGSGFSGSPSVTDPRPLNPSA---YSSTTLPA-ARAASPY-SQRPASPTAIR 286
 142 APGATVAAPRGSPK-OSPSRLAKSYSTSPINIVSSAGLSPIRVTSPTVQST 340
 143 JTSROT-SNENGPTPQQTARV-----GSPLTLTAQOT-----RVASPSQGV-V 333
 144 ?HQLSSTIGTYATLST-KRLVHASQYSKSOELVATATLQRPGLAAGSRASY 399
 145 ?---KRSQMTAVPQHLGFLQRTVHDMEQFQQQYDIYERVMPRPDLS-TGLRSSY 390
 146 ?HLPALQSPHEHIDPIYDRVYQKPPMRSLQSQGDPLPPAHTGYRTSTAP 459
 147 ?QLGQDLRSVSPDLHITPIYEGRTYYSVYRSPNHTVE-LQGSQALYRT---- 445
 148 ?DSV-PLQRTGSHQGNAAAATFORASVAGPASVADPYRQLOVCPSPVESYSK 518
 149 ?SGIGNLQRTSSOR-----STLTQRNVALNTATYAPYRQVQ-ECYNR 496
 150 ?PEGLTARSPSDTSIQDPRFGRWDPPELPEVIOHQFSPVQSNAAAYLOHLC 578
 151 ?PADDGTTSPSDTSIQDPRFGRWDPPELPEVIOHQFSPVQSNAAAYLOHLC 556
 152 ?KAEITROGGIOLLVLDHRTVHRACGALNLYVGKANDNKIALKXCGGI 638
 153 ?KMEVRLGGIKHLVDLDRVLEQVKNACGALNLYVGKANDNKIALKXCGGI 616
 154 ?LIRKTTDLRELVTVGLWNLSSCDALWPIQDALAVLTNAVITPHSGWNSPL 698
 155 ?LIRKSIDAEVRLTVGLWNLSSCDALWPIQDALAVLTNAVITPHSGWNSPL 676
 156 ?LQHSQVLRNATGLNRVSSAGEARRRMECDGLTDALLVIOALSGSEIDS 758
 157 ?LKFQTSVLNRNTGCLNLTSGAEARQMSCEGLVDSLLYVHTCVNTSDYDS 736
 158 ?ICVCIURLNLSYLAETSQGHMGTDELGLLCEANGKDAESSGCGWKKKKKK- 817

159 737 KTVENCVCITLRLNLSYRLEVPQARLLGLNELDLDLGKESPSKDSPTS-CWGKKK
 160 818 --SQDQWDGVGELPCDCAEPKGIOMLWHPISIVKPYLTLLSECSNPDTLEGAAGAL
 161 796 TPQEDQWDGVGPIGLSKSPKGVEMLWHPISIVKPYLTLLAESSNPATLEGSAGSL
 162 876 GSWKWSVVIIRAAVRKEKGLPILVELLRIDNRVVCATATARNMALDVRNKEELIGI
 163 856 SNWKAAYIRGGRPKRKGILVELLRIDNRVSSGATARNMALDVRNKEELIGI
 164 936 DLVHRLPGCNSNNTASKMSDDTAVCCCTLHEVITKNMENAKALRDAGGTEKL
 165 916 DLVNLPGNGPS-----VLSDETWAAICCALHEVITKNMENAKALDAGGTEKL
 166 996 SKGDKHSPKVVKAASQVLNSMWQYRDLRSLYKDGWSQYHFVASSSTIERDRQRP
 167 971 GRGRSSLKVVKAAQVLTNLQYRDLRSYKDGWQNHFTIPVSTILRDRFKSI
 168 1056 TSPISPV--RVSPNNRSASAPASPREMISLKERKTDYECTGSNATYHGAKEHTSI
 169 1027 -PSLSTNQMSPIIQSVGSSPALLGIRDRPRSEYDRTQPPMQYNSQGDATH
 170 1114 TAQNTGISTLYRNSYGAPAE-----IKHNOVSQOPVPOEPSRKDYETVQFQNSTI
 171 1084 YPGSKSPYIYSSYPARENRRLQHQQLYS--QDSSNRKNPDAYRLYLQSPH
 172 1170 SFEDQVHRRPPASEYTMHLGLSKTGNVDFYSAARPYSELNYETSHYPASPDSWI
 173 1142 PYFDRVHF-PASTDYSTQYGLKSTTVNDFYSTKRP-----SYRAEQYPGSPDSW
 174
 175 RESULT 4
 176 ARVC MOUSE STANDARD; PRT; 969 AA.
 177 ID ARVC MOUSE
 178 AC P98203;
 179 DT 16-OCT-2001 (Rel. 40, Created)
 180 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 181 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 182 DE Armadillo repeat protein deleted in velo-cardio-facial syndrome
 183 DE homolog (Fragment).
 184 GN ARVCF.
 185 OS Mus musculus (Mouse).
 186 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 187 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
 188 OX NCBI_TaxID=10090;
 189 RN [1]
 190 RP SEQUENCE FROM N.A.
 191 RX MEDLINE=20512094; PubMed=11058098;
 192 RA Kaufmann U., Zuppinger C., Waibler Z., Rudiger M., Urbich C.,
 193 RA Martin B., Jockusch B.M., Eppenberger H., Starzinski-Powitz A.;
 194 RT "The armadillo repeat region targets ARVCF to cadherin-based cell
 195 RT junctions."
 196 RL J. Cell Sci. 113:4121-4135(2000).
 197 CC - FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHEREN
 198 CC - SIMILARITY: (BY SIMILARITY)
 199 CC - SIMILARITY: Belongs to the beta-catenin family.
 200 CC - SIMILARITY: Contains 10 ARM repeats.
 201
 202 THIS SWISS-PROT entry is copyrighted. It is produced through a coll.
 203 between the Swiss Institute of Bioinformatics and the EMBL out-
 204 the European Bioinformatics Institute. There are no restriction.
 205 use by non-profit institutions as long as its content is in
 206 modified and this statement is not removed. Usage by and for c
 207 entities requires a license agreement (See http://www.isb-sib.ch/
 208 or send an email to license@isb-sib.ch).
 209
 210 EMBL; AJ243418; -; NOT_ANNOTATED_CDS.
 211 HSSP; P35222; IG3J.
 212 MGD; MGI:109620; Arvcf.
 213 GO; GO:0005737; C:cytoplasm; IDA.
 214 GO; GO:0005634; C:nucleus; IDA.
 215 GO; GO:0005886; C:plasma membrane; IDA.

```

734 QY 734 RECDGLTALLVYTSALGSSSEIDSKTVENCVCILRNLSYRLAAETSQGHGMGT
Db 536 RECEGLVDALLHALQSAVGRKDTONKSVENCVCIMRNLSYHVHKEVPDGADRYQEI
QY 792 GLLCGEANGKDAESSGGWGKKKKKKKQDQWD--GVGPLDPCAEPKGLIOMLWT
Db 596 GSTTQRRRKO--DASCPGGKAKGKKDAEMDNFDTLDLPKRTAAKGFELLY
QY 849 PYLTLLSECSNPDTLEGAGALQNLAAAGSKWSYIRAAVVRKEKGLPIILVELLRJ
Db 654 LYLISLLTESRNFNTLEAAAGALQNLASGNTWATYIRATVRKERGLPVLVELLQ
QY 909 VCAVATARNNALDVNRKELIGKYAMEDIYHRLPGGNNNSNTA--SKAMSDDTVT
Db 714 VRVAITARNLSLDQRNKDLIGSYAMTELVRNV--RNAQAPAHPSAHLEEDTV
QY 967 LHEVITKNMENAIALRDAGGIEKLVIGISKDGKHPKVKVKAASQVLSNMWYRI
Db 771 IHEIVSDSLDNARSLQARGVPALVALVAS--SQSVREAKAASHVLQTVWSYKI
QY 1027 KDGHSQVHFVASSSTTIERDRQRPSSSRTPSPV--RVSPNNRSASAPASPRE
Db 828 QRDGWTKSRF--QSASTATAGPKGTPSSGGFDOSTLPLVDKSLDGKSNTRDVTPLM
QY 1080 ---MISLKERXTDYECTGSNATVHGAKGHTSRKDMATQNTGISTLYRNSGAF
Db 887 GYATVDRERET---LGSNST-----GD-TSEKELLRVYGGQV-----YCGF

RESULT 5
ARVC_HUMAN
ID AC ARVC HUMAN STANDARD; PRT; 962 AA.
IC ARVC0192;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Armadillo repeat protein deleted in velo-cardio-facial syndrome.
GN ARVCF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP MEDLINE=97271559; PubMed=9126485;
RA Sirotkin H., O'Donnell H., DasGupta R., Halford S., St Jore B.,
RA Puech A., Parimoo S., Morrow B., Skoultschi A., Weisman S.,
RA Scambler P., Kucherlapati R.;
RT "Identification of a new human catenin gene family member (ARVCF
RT the region deleted in velo-cardio-facial syndrome.";
RL Genomics 41:75-83(1997).
CC -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERE
CC JUNCTIONS.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=000192-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=000192-2; Sequence=VSP 006739;
CC -!- TISSUE SPECIFICITY: FOUND IN ALL THE EXAMINED TISSUES INCLUD
CC HEART BRAIN, LIVER AND KIDNEY. FOUND AT LOW LEVEL IN LUNG.
CC -!- DISEASE: GENE DELETED IN VELO-CARDIO-FACIAL SYNDROME (VCFS) .
CC HEMIZYGOSITY IN ALL VCFS PATIENTS WITH INTERSTITIAL DELETIONS.
CC HEMIZYGOSITY MAY PLAY A ROLE IN THE ETIOLOGY OF SOME OF THE
CC PHENOTYPES ASSOCIATED WITH VCFS CHARACTERIZED BY A WIDE SPRC
CC PHENOTYPES, INCLUDING CONOTRUNCAL HEART DEFECTS, CLEFT PALAT
CC FACIAL DYSMORPHOLOGY.
CC -!- SIMILARITY: Belongs to the beta-catenin family.
CC -!- SIMILARITY: Contains 10 ARM repeats.

This SWISS-PROT entry is copyright. It is produced through a col-
between the Swiss Institute of Bioinformatics and the EMBL out-
the European Bioinformatics Institute. There are no restriction

```


: | | : | | : | : : |
 >SSPGVDSVFLORTGSQHGPQNAATAAATFORASYAAGPASNYADPYRQLQYCPSVES 514

[illegible]

all motility and cell scattering in response to
growth factor treatment (in vitro).
BINDS TO E-CADHERIN AT A JUXTAMEMBRANE SITE WITHIN THE
C DOMAIN. BINDS TO PRESENTIN-1 (BY SIMILARITY).
AR LOCATION: Adherens junction (By similarity).
! Contains at least 3 ARM repeats.
! Belongs to the beta-catenin family.

! Entry is copyright. It is produced through a collaboration
Swiss Institute of Bioinformatics and the EMBL outstation -
Bioinformatics Institute. There are no restrictions on its
profit institutions as long as its content is in no way
this statement is not removed. Usage by and for commercial
uses a license agreement (see <http://www.isb-sib.ch/announce/>
mailto:license@isb-sib.ch).

! ; BAA23384.1; -.
! 08938; ARM.
! 00225; Armadillo.
! ; Armadillo_seg; 2.
! ; ARM; 4.
! ; 176; ARM REPEAT; 2.
! ; Cytoskeleton; Structural protein;
! protein; Repeat.
! 1
! 20 59 ARM 1.
! 34 104 ARM 2.
! 20 162 ARM 3.
! 19 255 POLY-LYS.
! 34 264
! ; AA; 28928 MW; 065A886B8D2F58E0 CRC64;
! ; 21.5%; Score 1372; DB 1; Length 264;
! ; 100.0%; Pred.No.1.1e-58;
! ; conservative 0; Mismatches 0; Indels 0; Gaps 0;
! ;
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 622
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 60
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 682
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 120
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 742
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 180
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 802
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 240
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 826
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 264
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 834
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 838
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 842
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 846
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 850
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 854
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 858
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 862
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 866
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 870
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 874
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 878
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 882
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 886
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 890
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 894
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 898
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 902
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 906
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 910
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 914
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 918
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 922
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 926
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 930
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 934
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 938
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 942
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 946
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 950
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 954
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 958
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 962
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 966
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 970
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 974
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 978
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 982
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 986
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 990
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 994
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 998
! ;\$NAAAYLQHLRCGDNKIKAEIRROGGIQLLVLLDHRMTVEVRSACGALRNLYVG 1000

SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
TISSUE=Petal kidney;
MEDLINE=98317528; PubMed=9653641;
Keirsebilck A., Bonne S., Staes K., van Hengel J., Nolllet F.,
Reynolds A., van Roy F.;
"Molecular cloning of the human p120catn catenin gene (CTNND1):
expression of multiple alternatively spliced isoforms.";
Genomics 50:129-146(1998).
[2]
SEQUENCE FROM N.A. (ISOFORM 1AC).
TISSUE=Brain;
MEDLINE=97349984; PubMed=9205841;
Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
"Prediction of the coding sequences of unidentified human genes.
The complete sequences of 100 new cDNA clones from brain which ca
code for large proteins in vitro.";
DNA Res. 4:141-150(1997).
-!- FUNCTION: EFFICIENT TYROSINE KINASE SUBSTRATE IMPLICATED BOTH
CELL TRANSFORMATION BY SRC AND IN LIGAND-INDUCED RECEPTOR
SIGNALING THROUGH THE EGF, PDGF, CSF-1 AND EBBB2 RECEPTORS. T
ASSOCIATION OF CATELINS TO CADHERINS PRODUCES A COMPLEX WHICH
LINKED TO THE ACTIN FILAMENT NETWORK, AND WHICH SEEMS TO BE O
PRIMARY IMPORTANCE FOR CADHERINS CELL-ADHESION PROPERTIES.
-!- SUBUNIT: BELONGS TO A MULTIPROTEIN CELL-CELL ADHESION COMPLEX
ALSO CONTAINS E-CADHERIN, ALPHA-CATENIN, BETA-CATENIN, AND GA
CATENIN. BINDS TO PRESENTIN 1 C-TERMINAL FRAGMENT AND MUTUAL
COMPETES FOR E-CADHERIN.
-!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR IN PARTICULAR C
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=32;
Comment=Experimental confirmation may be lacking for some
isoforms;
Name=1ABC;
IsoId=O60716-1; Sequence=Displayed;
Name=1AB;
IsoId=O60716-2; Sequence=VSP_006743;
Name=1AC;
IsoId=O60716-3; Sequence=VSP_006745;
Name=1BC;
IsoId=O60716-4; Sequence=VSP_006744;
Name=1A;
IsoId=O60716-5; Sequence=VSP_006743, VSP_006745;
Name=1B;
IsoId=O60716-6; Sequence=VSP_006743, VSP_006744;
Name=1C;
IsoId=O60716-7; Sequence=VSP_006744, VSP_006745;
Name=1;
IsoId=O60716-8; Sequence=VSP_006743, VSP_006744, VSP_006745
Name=2ABC;
IsoId=O60716-9; Sequence=VSP_006740;
Name=2AB;
IsoId=O60716-10; Sequence=VSP_006740, VSP_006743;
Name=2AC;
IsoId=O60716-11; Sequence=VSP_006740, VSP_006745;
Name=2BC;
IsoId=O60716-12; Sequence=VSP_006740, VSP_006744;
Name=2A;
IsoId=O60716-13; Sequence=VSP_006740, VSP_006743, VSP_00674
Name=2B;
IsoId=O60716-14; Sequence=VSP_006740, VSP_006743, VSP_00674
Name=2C;
IsoId=O60716-15; Sequence=VSP_006740, VSP_006744, VSP_00674
Name=2;
IsoId=O60716-16; Sequence=VSP_006740, VSP_006743, VSP_00674
VSP_006745;
Name=3ABC;
IsoId=O60716-17; Sequence=VSP_006741;
Name=3AB;
IsoId=O60716-18; Sequence=VSP_006741, VSP_006743;
Name=3AC;
IsoId=O60716-19; Sequence=VSP_006741, VSP_006745;
Name=3BC;

ON WAS OBSERVED IN APPROXIMATELY 10% OF INVASIVE DUCTAL
CARCINOMAS INVESTIGATED.
TY: Belongs to the beta-catenin family.
TY: Contains 10 ARM repeats.

19. AAC39804.1; -
23. AAC39808.1; -
21. AAC39826.1; -
41. AAC39827.1; -
42. AAC39807.1; -
22. AAC39827.1; -
26. AAC39811.1; -
28. AAC39813.1; -
38. AAC39823.1; -
27. AAC39809.1; -
24. AAC39812.1; -
29. AAC39814.1; -
30. AAC39815.1; -
31. AAC39846.1; -
33. AAC39818.1; -
34. AAC39819.1; -
35. AAC39820.1; -
36. AAC39821.1; -
39. AAC39824.1; -
40. AAC39825.1; -
43. AAC39828.1; -
44. AAC39829.1; -
20. AAC39805.1; -
17. AAC39802.1; -
25. AAC39810.1; -
32. AAC39817.1; -
44. AAC39829.1; -
21. AAC39806.1; -
20. AAC39805.1; -
20. AAC39805.1; -
37. AAC39822.1; -
18. AAC39803.1; -
62. BAA200838.1; -
2515; CTNN01.

```

Query Match      21.5%; Score 1371; DB 1; Length 968;
Best Local Similarity 32.5%; Pred. No. 6.6e-58;
Matches 364; Conservative 169; Mismatches 335; Indels 252;

QY      41  DGSSETTTSAILASVKEQELFELTLELAEROIVASQLERCKLG-SETGSMSS
Db      2   DDSEVESTASILASVKEQEAQPEKLTALREERHVSQAQLERVRVSPQDANPLMA
QY     100  EGFQWQSDGQKQIEDLTTGLIWDSCIRSLQESGILDPDQYSTGERPSELLSQS
Db     62   RRHONGRFVGDALERQKFSDLK-----NG--PQDHSH-----LLYS
QY     160  SRPEGSFYPSAHSNQTALGETTPQLPARGTCABATQCSQSOGTTSRAGHLA
Db     103  QEPG-----QIVETYTEEDP--EGAMSVSVSETSDDGTRRT-----
QY     220  PPPPPPPREPAPSLGSAFHLPDAPPAALAAALYYSSSTLPPAPPGSGPLAAPQG
Db     138  -----
QY     280  LORGSAPEGATYAAPRGSSPKQSPRLAKSYSTSPINIVWSAGLSPIRVTSTP
Db     138  -----ETTVKKVKVTVTRTVQPVAMGPDGL-----
QY     340  TISSPFIHQLSST-----IGTYATLSPTKRLVHASQYKHSQELYAT
Db     169  SVSNYIYIQLGRDFKNGNGGPGYVQAGTATLPRNFHYPPDGYSRHVEDGY--
QY     388  PGSLAAGSRASYSQGHGLGP-----ELRALQSPPEHIDPIYEDRVYQKPPMRSL
Db     222  PGG-----SNYGSLSRVTRIEERYPSPMEGYRAPSROD-VYGPQPOVRVG

```

```

*HTGTYRTSTAPSPGVDSV-----PLQRTGSHQGPONAAAATFORASYAAG 494
*FHPEYGLDEDDQSRMGYDLDYGMMSDYGTARTGTP-----311
*VADPYROLQYCPSPVSPYSGSP-----ALPPEGTARSPSIDSIQK-DPRE 544
*-SDPRRLR---SYEDMTGEEVPSDQYWPAPLAQHERGSLA---SLDSLRKGGPPP 360
*DPPELPEVIMQ-CHOPSPSYQSNAAAYLOHLCFCGDNKIKAEIRROGGIOLLLDLHR 604
*QPELPEVIMLGLFLDAVKSNAAAYLOHLCYRNDKVKTDVRKLGIPVLVGLLDRP 420
*HRSACGALNLYGVKANDNDKIALKNCGGIPALVRLRLKTTDLREIRLVGTGLNWL 664
*HLGACGALKNTSFG-DOQNKIAIKNCDCGVPALVELLRKARDMDLITEVITGLNWL 479
*AKWPIIQDALAVITNAVIIIPHSWENSPLODDRKIQLHSSQVLRNATGCLRNVS 724
*SIKMEIVDHALHALTDEVIIIPHSWEREPNEDCKPRHIEWESVLTNTAGCLRNVS 539
*ARRMRBCDGLTDALLYVIQASGSSEIDSKTVENCVCILRNLSYRLAAETSQGH 784
*ARRKLRCDGLVDALIFIVQAEIGKDSKLVENCVCILRNLSYQVHREIPQ---596
*ELDGLLCEANGKDAESGCGKXKXKXKXKXQDQW--DVGGLPLPCA-----832
*RYQEAAPNVANNTGPHAASCFCGAKKKGK---DEWFSRGKKPIEDPANDTVDFFPKRT 650
*GIQWLHPSIVKPKYTLTLLSECSNPDTEGAGALONLAAGSKWSVYIRAAVRKEK 892
*GYELLFQPEVRIYISLLKSKTFAIIEASAGALONLCAGRTYGRVIRSAIRQEK 710
*LVELLIRDNDRVCAVATARNMALDVNRNELIGKYAMRDLVHRLPGGNNNSNTAS 952
*IADILLTNEHERVKAAGALRNLAVDARNKELIGKHAIPNLVXKLPQGGQN---SS 767
*ODTWTAVCCITLHEVITKMNENAKALRDAGGIEKLVIGSKGDKHSPKVKAASQV 1012
*EDTVISILNTINEVAENLEAAKRLRETOGIEKLVINKS--GNRSEKEVRAAALV 825
*QVQRDLRLSKYKKGWSYHFVASSSTIERDR-QRPYSSSRTPSTSPVRVSPNNRSA 1071
*QYKELRKLPEKEGKWSDFQVNLNNAASRQSSHSDSTLPLI-----DRNQKSD 880
3PREMISLKERKTDYECTGNSNATYHGAKGHTSRKD 1111
-DREEIQMNGSNKSLDNNYSPTPERGDNRTLD 918
STANDARD; PRT; 881 AA.
;
rel. 41, Created
rel. 41, Last sequence update)
rel. 41, Last annotation update)
(Human).
:azoa; Chordata; Craniata; Vertebrata; Euteleostomi;
:eria; Primates; Catarrhini; Hominidae; Homo.
36;
N.A. (ISOFORMS 1 AND 2).
101; PubMed=922383;
Jhn C., Franke W.W.;
2a and 2b: constitutive proteins of dual location in the
1 the desmosomal plaque."
135:1009-1025(1996).
[CITY.
194; PubMed=10374264;
Jhn C., Moll R., Schwetlick I., Franke W.W.;

```

```

RT "Desmosomal plakophilin 2 as a differentiation marker in normal
RT malignant tissues."
RL Differentiation 64:277-290(1999).
CC -!- FUNCTION: May play a role in junctional plaques.
CC -!- SUBCELLULAR LOCATION: Nuclear and associated with desmosomes
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=2; Synonyms=B;
CC IsoId=Q99959-1; Sequence=Displayed;
CC Name=1; Synonyms=A;
CC IsoId=Q99959-2; Sequence=VSP 006736;
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. FOUND AT DESMOSOMAL PLJ
CC IN SIMPLE AND STRATIFIED EPITHELIA AND IN NON-EPITHELIAL TISSU
CC SUCH AS MYOCARDIUM AND LYMPH NODE FOLLICLES. IN MOST STRATIF
CC EPITHELIA FOUND IN THE DESMOSOMES OF THE BASAL CELL LAYER AND
CC SEEMS TO BE ABSENT FROM SUPRABASAL STRATA.
CC -!- SIMILARITY: Belongs to the beta-catenin family.
CC -!- SIMILARITY: Contains 8 ARM repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a coll
CC between the Swiss Institute of Bioinformatics and the EMBL out
CC the European Bioinformatics Institute. There are no restriction
CC use by non-profit institutions as long as its content is ir
CC modified and this statement is not removed. Usage by and for c
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X97675; CAA6264.1; -
CC EMBL: X97675; CAA6265.1; -
CC Genew; HGNC:9024; PKP2.
CC MIN; 602861; -
CC GO; GO:0030057; C:desmosome; NAS.
CC GO; GO:0016021; C:integral to membrane; TAS.
CC GO; GO:0005634; C:nucleus; NAS.
CC GO; GO:0005886; C:plasma membrane; TAS.
CC GO; GO:0016337; P:cell-cell adhesion; NAS.
CC InterPro; IPR008938; ARM.
CC Pfam; PF00514; Armadillo_seg; 4.
CC SMART; SM00185; ARM; 4.
CC PROSITE; PSS0176; ARM_REPEAT; 1.
CC Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;
CC Repeat; Alternative splicing.
FT REPEAT 341 383 ARM 1.
FT REPEAT 385 424 ARM 2.
FT REPEAT 427 467 ARM 3.
FT REPEAT 571 616 ARM 4.
FT REPEAT 671 711 ARM 5.
FT REPEAT 719 758 ARM 6.
FT REPEAT 763 804 ARM 7.
FT REPEAT 807 849 ARM 8.
FT VARSPLIC 460 503 Missing (in isoform 1).
FT /FTID=VSP_006736.
SQ SEQUENCE 881 AA; 97398 MW; 947838B760BF1DSD CRC64;
Query Match 12.6%; Score 801; DB 1; Length 881;
Best Local Similarity 28.4%; Pred No. 6.3e-31;
Matches 259; Conservative 134; Mismatches 323; Indels 196; G
QY 256 SSTLPAPPGGSPLA--APQGGSPFK-----LQGGSAPEG-----ATYA
DB 27 SSSLALPSEAKLKLGGSGRGQTVKSLRIQEQVQQTARKGRSSVGNGLHRTSS
QY 297 -----GSSPKQSPSLAKSYSTSSPINIVWSAGLSPIRVTSPTVQS
DB 87 VYNLHLVENDFVGRSPVPKTYDMLKAGTTAT-----YEGRWGRGTAQYSQKSVEE
QY 344 SPTHQLSSITIGVATLSPTK-RLVHASEQVSK-----HSQELVATATLQRP--
DB 143 HPLRLLEISPD-----SPERATHYSDYQYSQSQAGHTLHHQESRAALLVPRY.
QY 391 LAAGSRASYSSQHGHLGPELRALQSPPEHIDPIYEDRVYQKP-----PMRSLSQSQ

```

VSAGTTSRQRHFDYHQYQ--HGSVSDTVFDSIPANPALLTYPRPFGTSRSMGNLL 256
 -----FLPPAHTCTYRTSTAPSSPGVDVSPLORTGSHQGPQNAATFQ 487
 NYLTAGLVGVQVPLVLPQVPT-QNRASRSSWHQSSFSHTLTLREAGPSVAVDSSGR 315
 YAAGPASNYADPYRQLQYCFVSVEPSYKSGPALPEGTTLARSPSIDSIQDKDPREFGW 547
 LTVGQAA-----AGSGNLLTERSTFTDS-----QLGN 346
 E--LPEVLQMLQ--HQPPSVQSNAAAYLOHLCFGDNKIKAEIRROGGIQLLVLLDH 603
 EMTLERAVSMLEADHMPSPSRISAATFIQHECFQKSEARKRVNQLRGILKLLQLVK 406
 EVHRSACALRNLYGKANDNDKIALKNCGGIPALVRLAKRTDLEIRELVT----- 658
 DVQRAVCGALRNLF--EDNDNKLEVAELNGVPELLQLVLAQTRDLETKKQITHTVYN 464
 -----GVLNLSLSCDALKMPIIQQDAL 679
 RNGWPGAVAHACNPSTLGGGGRITRSGVRDQDPQDQGLLNLSSNDKLNLMITEAL 524
 TNVLIIPHSWENSPLODDRKIQ-LHSSQVLRNATGCLRVSSAGEARRMRRECDS 738
 TENIILIPSGW---PEGDYPKANGLLDFIPYNTVTCGLRMSSAGADGRKAMRRCDS 581
 ALXYIQSALGSSIDSKTVCNLCILNLSYLAETS-----QGHWMGTDE 789
 SLVHYVGTIADYQDDKATENCVCILHNLISYQLEAELPEKYSQNIYIQRNQTQD- 640
 LLCGEANGKAESSGCKKKKKKQSDQWDVGPLPCABPPKGIQMLWHPISIVKP 849
 -----NNK---SIGCFGSRKVK--EQYQDV--PMPEEKNPKGVWELWHSIVIRM 685
 LLSGCSNPDTLEGAGALQNLAAQSKWKSYYI-RAAVRKEKGLPILVELLRIDNRY 908
 LIAKSVNYTQEAALGALQNLTAGSGEMPTSVAGTVVQKSGGLQHTPKMLHVGDPVS 745
 VATALRNALDVRNKLIGYAMRDLVHRLPGGNNNSNTASKAMSDDTVTAVCCTLH 968
 AISLLRNLSRLSLQNEIAKETPLDLVSIIP-----DTPVSTDLLIETTASACVTLN 800
 TKMENAKALRAGGIEKLVIKSKGDKX-SPKVVAASQVLNSWQYRDLSLYK 1027
 QNSYQNAARDLNTGGIQKIMAI--SAGDAYASKAKAASVLLSYLWHAHTLHAYK 858
 WSOYHVFVAS 1039
 FKKTDFVNS 870

STANDARD; PRT; 797 AA.

(Rel. 41, Created)

(Rel. 41, Last sequence update)

(Rel. 42, Last annotation update)

3.

(Human)

atazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Cheria; Primates; Catarrhini; Hominidae; Homo.
 506;

N.A.

carcinoma;

2495; PubMed-10374265;

Langbein D., Praetzel S., Rode M., Rackwitz H.-R.,

3 - a novel cell-type-specific desmosomal plaque
 ion 64:291-306(1999).

RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99310799; PubMed-10381383;
 RA Bonne S., van Hengel J., Nollet F., Kools P., van Roy F.;
 RT "Plakophilin-3, a novel armadillo-like protein present in nuclei
 RL desmosomes of epithelial cells."; J. Cell Sci. 112:2265-2276(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE-22388257; PubMed-12477932;
 RA Strausberg R.L., Feigold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk
 RA Villalon D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Faney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanche
 RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.C.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywicki M.I., Skaleka U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences."; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RL
 CC -!- FUNCTION: May play a role in junctional plaques.
 CC -!- SUBCELLULAR LOCATION: Nuclear and associated with desmosomes
 CC -!- TISSUE SPECIFICITY: FOUND IN DESMOSOMES OF MOST SIMPLE AND
 CC STRATIFIED EPITHELIA. NOT FOUND IN FORESKIN FIBROBLASTS AND
 CC CELLS OF LYMPHATIC FOLLICLES. BESIDE DENDRITIC RETICULA
 CC DESMOSOME-BEARING TISSUES.
 CC -!- SIMILARITY: Belongs to the beta-catenin family.
 CC -!- SIMILARITY: Contains 8 ARM repeats.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a col
 CC between the Swiss Institute of Bioinformatics and the EMBL ou
 CC the European Bioinformatics Institute. There are no restrictio
 CC use by non-profit institutions as long as its content is i
 CC modified and this statement is not removed. Usage by and for
 CC entities requires a license agreement (See http://www.isb-sib.ch
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z98265; CAB44310.1; -;
 DR EMBL; AF053719; AAF23050.1; -;
 DR EMBL; BC000081; AAH00081.1; -;
 DR Genew; HGNC:9025; PKP3.
 DR MIM; 605561; -;
 DR GO; GO:0005911; C:intercellular junction; TAS.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000225; Armadillo.
 DR Pfam; PF00514; Armadillo_seg; 2.
 DR SMART; SM00185; ARM; 3
 DR PROSITE; PS50176; ARM_REPEAT; 1.
 KW Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein
 FT REPEAT. 305 348 ARM 1.
 FT REPEAT. 351 390 ARM 2.
 FT REPEAT. 393 432 ARM 3.
 FT REPEAT. 449 487 ARM 4.
 FT REPEAT. 491 536 ARM 5.
 FT REPEAT. 596 637 ARM 6.
 FT REPEAT. 645 684 ARM 7.
 FT REPEAT. 689 730 ARM 8.
 SQ SEQUENCE 797 AA; 87081 MW; D43C7E77FA805E7E CRC64;

Query Match

12.1%; Score 770.5; DB 1; Length 797;


```

-----LPDVRGLDSTYG-----HRTIQLRSS-----GFDDIDLPsAVKYLMA 330
SVQNSAAAYLOHLFCFGDNKKIKABIRROGGIOLVLDLDRHMTVEHRSAGCALRNLVY 621
NLQVLGAAYI0HRCYSYDAAAKQARSLOAVPRLVKLFNFHANEQVORHATGAMRNLIY 390
NDDNKIALKNCGGIPALVRLLRKTTDLREILVTGVLNLSSCDALKMPITODALAV 681
NVDNKLALVENGIFELLRTLRQDD--ELRKNVTGILNLSSSHLLKORLARDLTLEQ 447
AVIIPHSGWENSPLODDRKKIOLHSS--QVLRNATGCLNENVSAGEEARRRRMRECDGL 739
LVLSPLSGPGGPPL-----IQONASEABIFYNATGFLNLSASQATQKMKRECHGL 502
LLYVIQALSGSSEIDSKTVENCVCILRNLSYRL-----AAFTSQOHHMTDELDG 792
LVTVINHALDVGKCEDKSVENAVCVLRNLSYRLDEWPPSALQRLLEGH--GRRDMAG 560
GEANGKDAESSGCGKKKKKKKSQDQWDGVGPLPCDAEPPKGIOMLWHPSTVVKPYLT 853
G-----EMVGCFTPQSRRLRELPLTADALTFAEVSKDPKGLEWLSFQIVGLYNR 613
ECS-NPTLEGAAGALONLAGSCKWS--VYTRAARVREKKGPIVLVELLRIDNRVVC 910
RCELNRHRTTEAAGALONITAGDRRWAGVLSRLALEQERILNPLIDRVRTADHNQLR 673
TALFNMALDVENKELIGKYAMRDLVHRLPGGNSNNITASKAMSDDTVTAVCCTLHEV 970
GLIENLSRNRKNDKMTKVVSHLEKLPD-----SVGEKCPDPAEVLVNIITAVLNLL 728
NMENAKALRDAGGIEKLVGISKSGDKHSPKVKKAASQVLNMSMQYRDLRSLYKKG 1030
SPTAARDLLYFDGLRKLVLVFKKRDSPDSEKSSRAASSLLANLWQYSKLHRDFRAKG 788
YHFV 1037
EDFL 795

STANDARD; PRT; 728 AA.

(Rel. 4i, Created)
(Rel. 4i, Last sequence update)
(Rel. 4i, Last annotation update)
1.

(Mouse).
etazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Theria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
0090;

M N.A.
/6: TISSUE=Skin;
Hunziker A.H., Franke W.W.;
EP1996) to the EMBL/GenBank/DBJ databases.
: SEEMS TO PLAY A ROLE IN JUNCTIONAL PLAQUES (BY
TY).
LAR LOCATION: Nuclear and associated with desmosomes (By
ty).
TY: Belongs to the beta-catenin family.
TY: Contains 7 ARM repeats.

ROT entry is copyright. It is produced through a collaboration
Swiss Institute of Bioinformatics and the EMBL outstation -
Bioinformatics Institute. There are no restrictions on its
-profit institutions as long as its content is in no way
this statement is not removed. Usage by and for commercial/
quires a license agreement (See http://www.isb-sib.ch/announcement/
mail to license@isb-sib.ch).

CAA69240.1? -

```


YTVRLMTSQPMQAKQYFNSMLNVNLCRNTAS--SPKAAEAAARLLLSMDWAS 695

3LYKKDQWSQ 1033

1 : : : :

3VLRQGGFDR 709

STANDARD; PRT; 747 AA.

1; Q15152;

rel. 41, Created)

rel. 41, Last sequence update)

rel. 41, Last annotation update)

(Band-6-protein) (B6P).

(Human).

taxon: Chordata; Craniata; Vertebrata; Euteleostomi;

Phyla: Primates; Catarrhini; Hominoidea; Homo.

16;

N.A. (ISOFORM 1).

16; PubMed=7527055;

Kristjansson G.I., Plessmann U., Weber K.;

in, a major constituent of desmosomes from stratified

a novel member of the armadillo multigene family.;

107:2259-2270(1994).

N.A. (ISOFORM 1).

1994) to the EMBL/GenBank/DBJ databases.

N.A. (ISOFORMS 1 AND 2).

172; PubMed=9369526;

ingbein L., Rode M., Praetzel S., Zimbelmann R.,

la and 1b: widespread nuclear proteins recruited in

cellular cells as desmosomal plaque components.;

as. 290:481-499(1997).

SEEMS TO PLAY A ROLE IN JUNCTIONAL PLAQUES.

AR LOCATION: Nuclear. Isoform 1 is also associated with

1.

FE PRODUCTS:

alternative splicing; Named isoforms=2;

nonyma=B;

13835-1; Sequence=Displayed;

nonyma=A;

13835-2; Sequence=VSP_006735;

SPECIFICITY: NUCLEAR ISOFORM IS WIDELY EXPRESSED. ISOFORM 1

13835-1; Sequence=Displayed;

DEFECTS IN PKP1 ARE ASSOCIATED WITH ECTODERMAL

'SKIN FRAGILITY SYNDROME.

1; Belongs to the beta-catenin family.

1; Contains 8 ARM repeats.

1; entry is copyright. It is produced through a collaboration

Swiss Institute of Bioinformatics and the EMBL outstation -

profit institutions as long as its content is in no way

this statement is not removed. Usage by and for commercial

ires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)

ail to license@isb-sib.ch).

CAAS5981.1; -

CAAB4426.1; -

-; NOT ANNOTATED_CDS.

CAAB4426.1; -

223; PKP1.

-

-

7; C:desmosome; NAS.

DR GO: 0005634; C:nucleus; NAS.

DR GO: 0019215; F:intermediate filament binding; NAS.

DR GO: 0030280; P:structural constituent of epidermis; NAS.

DR GO: 0007155; P:cell adhesion; NAS.

DR GO: 0007165; P:signal transduction; NAS.

DR InterPro: IPR008938; ARM.

DR InterPro: IPR000225; Armadillo.

DR Pfam: PF00514; Armadillo_seg; 4.

DR SMART: SM00185; ARM; 5.

DR PROSITE: PS00176; ARM_REPEAT; 3.

KW Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;

KW Repeat; Alternative splicing.

FT REPEAT 234 274 ARM 1.

FT REPEAT 277 316 ARM 2.

FT REPEAT 319 359 ARM 3.

FT REPEAT 439 484 ARM 4.

FT REPEAT 537 577 ARM 5.

FT REPEAT 585 624 ARM 6.

FT REPEAT 626 670 ARM 7.

FT REPEAT 673 713 ARM 8.

FT VARSPLIC 412 432

Missing (in isoform 1).

/FTID=VSP_006735.

FT CONFLICT 154 154

R -> G (IN REF. 1).

FT CONFLICT 216 222

PPISNK -> RHLLQ (IN REF. 1).

FT CONFLICT 462 462

V -> E (IN REF. 1).

FT CONFLICT 496 496

Q -> K (IN REF. 1).

FT CONFLICT 506 506

T -> P (IN REF. 1).

FT CONFLICT 553 553

L -> S (IN REF. 1).

SQ SEQUENCE 747 AA; 82860 MW; 60C1BC50AB4B6F CRC64;

Query Match 11.0%; Score 699; DB 1; Length 747;

Best Local Similarity 29.0%; Pred. No. 3.5e-26;

Matches 219; Conservative 108; Mismatches 271; Indels 156; G

Qy 365 LVHASTQYKHSQELATATLQRPGLAAG-----SRASYSQ-----HGHL

Db 46 MMTVKRQKSKSQS--STLSHNRGMYDGLADNYYNITTSRSYYSKFOAGNSW

Qy 411 RALQSPHHIDPIEDRVYQKPPMRSLSQSQGDPPLPAHTGYRTSTAPSSPGVDS

Db 101 -----YPIYNGTLKREPDRRFS-----S

Qy 471 RTGSHGPNQAAAT-----FORASYAAGPASYADPYRLQY-----CPS

Db 124 ENWSRHYPRGSCMTTGAGSDICPMQIKASRSPDLVCDPRGLTKGTIGSKGQKT

Qy 516 YS-----KSGPALPPEGLARSPSI-----DSIQKDPREGW-----R

Db 184 YSPYSTCGGKATKCFVRPPSCASKQDPVYIPISCNKD-LSFGHSRASSKICSE

Qy 552 ----LPEVIQMLQHPFVSQSNAAVYLQHLFCGDNKIKAEIRQGGIQLLVLDLH

Db 243 SGTIPKAVQYLSQDEKYAIGAYVYIQTCTODESAKQOVYQIGGICKLVLLRS

Qy 608 VHSACALRNLYGKANDDNKIALKNCGGIPALVRLRLKRTDLEIRLVGLWN

Db 303 VQQAAGALRNLYF--RSTTNKLETRQNGIREAVSLRRRTGNAETQKLTGLLN

Qy 668 DALKMPILQDALAVLNNAVIIIPHSCHENSPLQDRKILQLHSSQVLNATCIR---

Db 361 DELKEELIADALPVLADRVIIIPFSGMCDGNNSMREV--DPEVFNATCLRKRL

Qy 721 -----NVSSAGEEARRMRECDGLTDALLVVIQSLAGSSEIDSXTVEI

Db 419 LLALVFORATSSRVNLSSA-DAGROTWRNYSGLIDSLMAYVQNCVAASRCDKSVI

Qy 767 ILRLSYRLAAETSSQGHMGTDELGLLCEANGKDA-----ESSGCGWKKKKKKKS

Db 478 VLHNLVRLDAEVPTRYQ-----LEYNARNAYTEKSTGCFNSKSDKMMN

Qy 823 DGVGPLPDCAEPKGIQMLWHPISIVPKPYLTLLSECSNPDTLEGAAGALQNLASGW

Db 527 D--CPLPEETNPKSGWLYHSDAIRTYLNLGMGSKKDATLEACAGALQNLTSKG

STANDARD; PRT; 727 AA.

(Rel. 41, Created)
(Rel. 41, Last sequence update)
(Rel. 41, Last annotation update)
1 (Band-6-protein) (B6P).

Bovine).
etazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
theria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
inae; Bos.
913;

M N.A.
6971; PubMed=7890138;
chmidt A., Zimbelmann R., Schaefer S.,
owski S., Stumm S., Keith M., Figge U., Schnolzer M.,
specific desmosomal plaque proteins of the plakoglobin
ophillin 1 (band 6 protein).",
ion 58:113-131(1994).
; SEEMS TO PLAY A ROLE IN JUNCTIONAL PLAQUES (BY
TY).

LAR LOCATION: Nuclear and associated with desmosomes (By
ty).
TY: Belongs to the beta-catenin family.
TY: Contains 8 ARM repeats.

3	CAA86029.1; -	
57	C; desmosome; ISS.	
34	C; nucleus; ISS.	
15	F; intermediate filament binding; ISS.	
80	F; structural constituent of epidermis; ISS.	
55	P; cell adhesion; ISS.	
65	P; signal transduction; ISS.	
	RO08938; ARM.	
	RO00225; Armadillo.	
4	Armadillo_seg; 4.	
85	ARM; 5.	
0176	ARM_REPEAT; 3.	
n	Cytoskeleton; Structural protein; Nuclear protein;	
235	275	ARM 1.
278	317	ARM 2.
320	360	ARM 3.
419	464	ARM 4.
517	557	ARM 5.
565	604	ARM 6.
606	650	ARM 7.

```

RESULT 15
DREL HUMAN
ID DRPL HUMAN STANDARD; PRT; 1185 AA.
AC AC P54259; Q99495; Q99621; Q9UEK7;
DT 01-OCT-1996 (Rel. 34, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Atrophin-1 (pentacubital-pallidolysian atrophy protein).
GN DRPLA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Brain, and Cerebellum;
RC MEDLINE=95144175; PubMed=7842016;

```

B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	B16	B17	B18	B19	B20	B21	B22	B23	B24	B25	B26	B27	B28	B29	B30	B31	B32	B33	B34	B35	B36	B37	B38	B39	B40	B41	B42	B43	B44	B45	B46	B47	B48	B49	B50	B51	B52	B53	B54	B55	B56	B57	B58	B59	B60	B61	B62	B63	B64	B65	B66	B67	B68	B69	B70	B71	B72	B73	B74	B75	B76	B77	B78	B79	B80	B81	B82	B83	B84	B85	B86	B87	B88	B89	B90	B91	B92	B93	B94	B95	B96	B97	B98	B99	B100
----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 CompuGen Ltd.
 in search, using sw model
 sil 22, 2004, 12:10:15 ; Search time 57 Seconds
 (without alignments)
 6780.874 Million cell updates/sec

-09-501-171a-4
 77
 4FARKPPGAAPLAMPVPDQ.....PYSELNYETSHYPASPDSDWV 1225
)SUM62
 op 10.0 , Gapext 0.5
 .7041 seqs, 315518202 residues
 is satisfying chosen parameters: 1017041
 yth: 0
 fth: 2000000000
 inimum Match 0%
 ximum Match 100%
 isting first 45 summaries

PREMBL 25: *
 sp_archaea: *
 sp_bacteria: *
 sp_fungi: *
 sp_human: *
 sp_invertebrate: *
 sp_mammal: *
 sp_mhc: *
 sp_organelle: *
 sp_phage: *
 sp_plant: *
 sp_rodent: *
 sp_virus: *
 sp_vertebrate: *
 sp_unclassified: *
 sp_virus: *
 sp_bacteriapi: *
 sp_archaeap: *

the number of results predicted by chance to have a
 : than or equal to the score of the result being printed,
 id by analysis of the total score distribution.

SUMMARIES

Seq	Length	DB	ID	Description
1	1265	13	Q801N7	Q801n7 xenopus lae
2	1149	4	Q86W91	Q86w91 homo sapien
4	1168	13	Q8AXM8	Q8axm8 xenopus lae
4	983	13	Q7T071	Q7t071 gallus gall
1	907	13	Q9DFB2	Q9dfb2 xenopus lae
9	742	13	Q9DE61	Q9de61 xenopus lae
3	956	11	Q7TS93	Q7ts93 mus musculus
2	956	11	Q924L4	Q924l4 mus musculus
1	956	11	Q8C3U7	Q8c3u7 mus musculus
1	962	11	Q924L5	Q924l5 mus musculus
8	938	11	Q80XQ4	Q80xq4 mus musculus
7	945	11	Q8CHF8	Q8chf8 mus musculus
2	295	13	Q90X45	Q90x45 brachydanio
9	892	11	Q8BRF2	Q8brf2 mus musculus
9	892	11	Q924L2	Q924l2 mus musculus
8	898	11	Q8BQ36	Q8bq36 mus musculus

ID	Q801N7	PRELIMINARY;	PRT; 1265 AA.
AC	Q801N7;		
DT	01-JUN-2003 (TREMELrel. 24, Created)		
DT	01-JUN-2003 (TREMELrel. 24, Last sequence update)		
DT	01-OCT-2003 (TREMELrel. 25, Last annotation update)		
DE	Similar to plakophilin 4 (Fragment).		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI_TaxID=8355;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Embryo;		
RA	Klein S., Strausberg R.;		
RL	Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BCO48023; AAH48023.1; -		
DR	InterPro; IPR008938; ARM.		
DR	InterPro; IPR000225; Armadillo.		
DR	Pfam; PF00514; Armadillo_seg; 4.		
DR	SMART; SM00185; ARM; 6.		
DR	PROSITE; PS50176; ARM_REPEAT; 3.		
FT	NON_TER 1		
SQ	SEQUENCE 1265 AA; 139253 MW; 01F7CB67BBAAE2FB CRC64;		

ALIGNMENTS

RESULT 1

Q801N7	Q801N7	PRELIMINARY;	PRT; 1265 AA.
AC	Q801N7;		
DT	01-JUN-2003 (TREMELrel. 24, Created)		
DT	01-JUN-2003 (TREMELrel. 24, Last sequence update)		
DT	01-OCT-2003 (TREMELrel. 25, Last annotation update)		
DE	Similar to plakophilin 4 (Fragment).		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI_TaxID=8355;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Embryo;		
RA	Klein S., Strausberg R.;		
RL	Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BCO48023; AAH48023.1; -		
DR	InterPro; IPR008938; ARM.		
DR	InterPro; IPR000225; Armadillo.		
DR	Pfam; PF00514; Armadillo_seg; 4.		
DR	SMART; SM00185; ARM; 6.		
DR	PROSITE; PS50176; ARM_REPEAT; 3.		
FT	NON_TER 1		
SQ	SEQUENCE 1265 AA; 139253 MW; 01F7CB67BBAAE2FB CRC64;		

Query Match 44.6%; Score 2842.5; DB 13; Length 1265;
 Best Local Similarity 48.4%; Pred. No. 1.1e-173;
 Matches 622; Conservative 192; Mismatches 319; Indels 153; G:

QY	11	PLGAMPVPQPS	---SASEKTSLSPLNTSGDSETTTSA-ILASVKEQLQFI
Db	62	PAGGVPVPGQTSLMEALQITQAAPS-NT----	GMEPETTATNIIASVKEQLQFI
QY	67	RELEAFRQIVASQLERCKLGSETGSMSSAAEFQFQWQSDQCKDIEELTTGLEI	
Db	117	RELEVERQIVANQLERCKRUGAESPSIASVSSTKSPFWRK----	GDVPAGGSPKPRI


```

>PEGTIARSPSIDSIQKDPREFGWRDPELPEVIOHLOHQPSPVQSNAAAYLOHLCF 579
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>ADGGTTRSPSIDSIQKDPREFAWRDPPELPEVIMHMQHQPSPVQANAAAYLOHLCF 557
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>KAEIRQGGIQLVLLDHRMTEVHRSACGALRNLYVKGANDNKIALKNCGGIP 639
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>KMEVCRLLGGIKHLVDLDRHVEVOKNACGALRNLYVFGKSTDENKIAMQVGGIP 617
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>LRKTTDLRELVTVGLVNLSSCDALMKPIIQDALAVLTNAVIIIPHSWENSPQLQ 699
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>LRKSIDAEVRELTVGLVNLSSCDVAKVTIIRDALSTLTNTVIVHSGWNNSSFD 677
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>QLHSSQVLNATGCLRNVSAGBEARRRMRECDGLTDALLYVIOQALSGSEIDSK 759
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>KFQTSVLNRNTGCLRNLSAGAEARKQMRSCGELVDLSLLYVHTCVNTSDYDSK 737
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>VCILNLSVRLAAETSOQGMGTDELGLLGCFAANGKDAESSGCGWKKKKKK-- 817
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>VCTLRNLSVRLLEVQAKLLGLDELDDLLGKESPKOSEPS-CWGKKKKKKRT 796
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>WDVGGLPDCABPPKGIQMLHPSIVKPYLTLLSECSNPDTLLEGAGALQNLAAQ 876
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>WDVGGLPGLSKSPKGVEMLWHPVVKPYLTLLAESSNPATLEGSAGSIQNLASG 856
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>VYIRAAVRKEKGLPILVELLRIDNDRVVCVATARNMALDVRNKLIGKYAMRD 936
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>VYIRAAVRKEKGLPILVELLRIDNDRVVCVATARNMALDVRNKLIGKYAMRD 916
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>PGNNNSNTASKAMDDTVTAVCTTLHEVITKNMENAKALRDAGGIEKLVGLSKS 996
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>PGNGGPS-----VLSDETAAALCCALHEVTSKNMENAKALADSGGIEKLVNITKG 971
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>ISPKVYKAAASQVLNLSMWYDRLSLYKDGWSQYHFVASSSTLERDRQRPSSRT 1056
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>ISLKVYKAAQVLTNLQYRLRSIYKDGWNQNHITPVSTLERDRFXSH----- 1026
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>V--RVSPNRSASAPSPREMISLKERKTDYECTGSNATYHGAKGHTSRKDAWT 1114
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>TNQMSPIIQSGSKSP----- 1049
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>ISTLYRNSYGAPEAD-----IKHNOVSAQVPQEPSRKDYETVQPFQNSTRYDES 1170
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>-----IVISSYSPARQNRRLQHQLYYS--QDSSNRKNFDAYRLVQSPHSTEDP 1099
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>VHHRPPASEYTMHLGLKSTGNVDFYSAARPYSELNYETSHYPASPDSMW 1225
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
>VHF-PASTDYSYQYGLKSTNTVDFYSTKRP-----SYRAEQYGPSDSMW 1149
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
ELIMINARY; PRT; 1168 AA.

>EMBLrel. 23, Created)
>EMBLrel. 23, Last sequence update)
>EMBLrel. 25, Last annotation update)

; (African clawed frog).
; azoa; Chordata; Craniata; Vertebrata; Euteleostomi;
; achia; Anura; Mesobatrachia; Pipidea; Pipidae;
; enopus.
; 5;
; N.A.
; Vanlandschoot A., Staes K., van Roy F.;
; ated proteins in Xenopus laevis.";
; (-1999) to the EMBL/GenBank/DBS databases.
; ; AA013695.1; -
; 06938; ARM.
; 00225; Armadillo.
; Armadillo seg; 5-
; 76; ARM_REPEAT; 3.
; 8 AA; 128678 MW; 229DF17F55AA760D CRC64;

```

```

Query Match 43.4%; Score 2769.5; DB 13; Length 1168;
Best Local Similarity 48.6%; Pred. No. 4.9e-169;
Matches 609; Conservative 180; Mismatches 338; Indels 125; G

QY 15 MPVPDPSSASRKTSLSPGLNTSNGDSGTETTSAILASVKEQELFERLTRELE
Db 1 MPVPEQTPLMEERALKITQSAPSTGTMEPTTATN-ILASVKEQELFERLTRELE
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 75 IVASOLERCKLGSSETSMSSMAEFOFQWQSQDQKXIDELTTGLELYDS----
Db 60 IVANQLERCKLGAESPSIGSVSTEXSPWRS-----GDVPAGGPKPRLSDSHPN
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 129 RSLQE-SGILDPQDYSTGERPSLL--SOSALQLNSKPEGSFOYPASVHSNOTLALG
Db 116 RSEPEPMNAVYSPQSSILHREGSMGNSRSQTOMNSYSDSGYQAGSFHNN--IG
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 185 PSQLPARGT-----QARATGQSPSQG-----TTSRAGHLAGEPEAPPPPP
Db 173 RAQLSFTGSGTGNHAVRNSRAEGQTFVQPPNTMAGSRAVRRVSVSPRSQSPSYVI
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 225 -PPREPPAPSLGSAFHLDPAPPAALAAALYSSSTLPAPPRGSPPLAAPQGGSPTK
Db 233 SPBGSRLRTSVGSGYSPSTSDPRPLATNYSSTLLPS-QRAASPYTLHRSNPTV
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 284 GSAPEGATYAAPRGSPKQSPKSLAKSYST-----SSPINIVVSSAGLSPIRVT
Db 292 GSI-----NSRQLSNPNTWTSYQTLGSSARVGSPL---TSSDAHSRVGLS
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 337 VQTTISSSPIHQLSSTIGTYATLSPTKRLVHASEQYKHSQELYATATLORPQSLA
Db 336 -QAPMGSSP-----KRAGMTAVPQHTGTL-NDIDQEQQHYDIYERMVPRPDSL-
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 397 ASYSSOHGHLGPELRALQSPERHIDPIYEDRVYQKPPMRSLSSQSGDPLPAAHTGT
Db 388 SSTASQHSQJLQELRSVAVSDLOITPIYEGRTYSPVYRNANHAADLRQGSHTAI
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 457 TAPSSPGVDSVPLQRTGSHQGPQNAATAATFORASYAAGPASNADPYRQYQPCPV
Db 447 -----GSGSGNLQRTMSQR-----CTLAQRNNYALNTAANYAEPVRSQYRIS-
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 517 SKSGPALPEGLTLARSPSIDSIQKDPREFGWRDPELPEVIOHLOHQPSPVQSNAAA
Db 495 NRIPTHTPPDDGATRSFSDSIQKDPREFAWRDPPELPEVIMHMQHQPSPVQANAAA
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 577 LCFGDNKIKAETRRQGGIOLLVDLDRHMTVEHRSACGALRNLYVKGANDNKIAL
Db 555 LCFGDHKTLEVCRMGGTKHLVDLDRHMTVEHRSACGALRNLYVKGANDNKIAL
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 637 GIPALVRLRKTTDLRELVTVGLVNLSSCDALMKPIIQDALAVLTNAVIIIPHS
Db 615 GVPALLRLKKTSDQEVRELTVGLVNLSSCDVAKVTIIRDALSTLTNTVIVHSG
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 697 PLODDRKIQLHSSQVLNATGCLRNVSAGBEARRRMRECDGLTDALLYVIOQALG.
Db 675 SFDDDHKMKFQSSVLRNTTTCMRNLSSSGEAEARKQMRSCGELVDLSLLFYIQTGVN
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 757 DSKTVENCVCILRNLSVRLAAETSQCHMGTDLDGLLGCFAANGKDAESSGCGWKKK
Db 735 DSKTVENCVCILRNLSVRLLEVQAKLLGLDELDDLLGKESPKOSEPS-CWGKKK
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 817 K---SQDQWDGVGGLPDCABPPKGIQMLHPSIVKPYLTLLSECSNPDTLEGAAGA.
Db 793 KKKPPQEDLWDGVGPIPGLSQSPKGVEMLWHPVVKPYLTLLAESSNPATLEGSAGSI
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 874 AGSWKWSYVIRAAVVRKEKGLPILVELLRIDNDRVVCVATARNMALDVRNKLIG
Db 853 SAGNWKFAAYIRAAVVRKEKGLPILVELLRIDNDRVVCVATARNMALDVRNKLIG
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 934 MRDLVHRLPGGNNSNTASKAMDDTVTAVCTTLHEVITKNMENAKALRDAGGIEKI
Db 913 MRDLVNLPLPGNGGPS-----ILSDDTVAAALCCALHEVTSKNMENAKALADTGGIEKI
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Qy	502	PYQLQYCPSVES-----PY---SKGPALPPE-GTLA-----RSPSIDS
Db	324	PHKGRGYDDPIETEMVEERIPYLHGSYAAPLAQPERGSMASIDRLGKESPSIDS
Qy	544	EFGWRDPPELPEVLOMLQHOFPVSQNSAAAYLOHLCFGDNKIKAEIRROGGTQLL
Db	384	--WRDPDPEVJAMLSHDPVKNSAAAYLOHLCYENDKIKKQVRHUKGPIL
Qy	604	RMTVEHRSACGALRNLVYGKANDDNKIALKNCGGIPALVRLLRKTTDLLEIRELV
Db	441	PKPEVHRKACGALRNISYKDN-ENKVAIKNCDGIPALIRLLRKTNDMEARELY
Qy	664	LSSCDALKMPTIODALAVLTNAVITPHSGWENSPLQDDRKIQLHSSQVLRNATGI
Db	500	LSSEYELKMWIINHGLQTLUTNEVIIPHSGWENSEPNEKSPRDAEWTTVFKNTSG
Qy	724	SAGEEARRRMRECDGLTDALLYIOSALGSSEIDSKTVENVCILRNLSYLAAM
Db	560	SDGAERRELRCDGLVDALLHALQSAVGKDDTKNSVENCVCIMRNLSYHVHKK
Qy	784	HMGTDLDGLLGEANG-----KDAESSGCGWKKKKKKKQDQWD---GVGPILI
Db	617	--GADYQBELDAGQTAGTGGSKKKKKDAGCFGGKKAKGKNGGLDRSFDLILI
Qy	835	PKGIQMLWHPSTVKPYLTILSSCSNPDLTLEAGAGALQNLAAAGSKWKSYYIARAIV
Db	675	AKGFELLYQPDVVRLYLSILITSEONFNTLEAAAGALQNLGAGNTWNTYTRATVI
Qy	895	PILVELLRIDNRVVCATARNMALDVRNKEILIGKYAMRDLVHRLPGGNNNSNI
Db	735	PVLVELLQSDSKVVRAYSIALRNLSMDRRNKDILIGSAMGELVRNLP---SROC
Qy	955	MSDDTVTACCTTBEVITKWNENAKLRDAGGIEKLVGISKSKGDKHSPKVVAKAF
Db	792	LEEDTVAVLNTIHEIITDSSENARSLIQTGGIOKLVAISKS---SOSPRTKAF
Qy	1015	SNWQYDRDLRSLYKKGWQSYHFVASSSTIERDR---QRPYSSSRKTPSISPVRYE
Db	849	MTWSYKELRNALQKDGWNKSHFQSVSATPKSKSGKTAGRSGYDDSTLPLVD--KSC
Qy	1071	ASAPASPREMISL-----KERKTYECTGNSATYHGAKGEHTSRK
Db	907	GS-----RDMIPMDLPGDGYSTDHRDKERYKTSYDNTGDAKEPLKNDTNRK
Qy	1116	QNTGISTLYRNSYGAPAEIDIKHNQVSAQPV 1145
Db	962	NRASYNLV-----DARDIK-----PopV 979

```

RESULT 5
Q9DFB2
ID C9DFB2 PRELIMINARY; PRT; 907 AA.
AC Q9DFB2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Catenin arvcf-2ABC protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
NCBI_TaxID=8355;
[1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=20459074; PubMed=10899158;
RA Paulson A.F., Mooney E., Fang X., Ji H., McCrea P.D.;
RT "Xarvcf; Xenopus Member of the p120 Catenin Subfamily Association
RT Cadherin Juxtamembrane Region.";
RL J. Biol. Chem. 275:30124-30131 (2000).
DR EMBL; AF287051; AAG00555.1; -.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.

```


[illegible]

```

Mouse).
azoa; Chordata; Craniata; Vertebrata; Euteleostomi;
eria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
90;
N.A.
; Puech A., Merscher S., Xu H., Kucherlapati R.,
- expression analysis of Arvcf, a candidate gene for
cial syndrome.";
--2000) to the EMBL/GenBank/DBJ databases.
; AAK64215.1; -.
0; Arvcf.
; C:cytoplasm; IDA.
; C:nucleus; IDA.
; C:plasma membrane; IDA.
; F:protein binding; IPI.
08938; ARM.
00225; Armadillo.
; Armadillo_seg; 4.
; ARM; 5.
; 76; ARM_REPEAT; 3.
; AA; 104165 MW; 19E9B3F33F346106 CRC64;
22.8%; Score 1417; DB 11; Length 956;
rity 34.8%; Pred.No.3.1e-82;
nservative 147; Mismatches 330; Indels 250; Gaps 39;
LASVKEQELQERLRELEAREQIVASQERCKL-GSETGSMSSSAEEQFQWQ 105
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
LASVKEQEARFERLTRALEQERHVALQLEAQQPFGSMGVMVG-SGQPLPMAWQ 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XOIEDELATGLVDSLCIRSLQESGILPDQDYSTGCRPSLLSQ-----SALQNSK 161
-----LVLQEQSPGSOASLATMPAPEVLEETVT 96
QVPASYHNQTLALGETTSPQLPARGTOARATGQFSQGTTSRAGHL-AGPEPAP 220
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XGPTSHSVTSVSEDGTT-----RRTETKVT-KTVKVTTRTVRQVPLPGDGLP 148
PPPPPPPPFAPSLGASAPHLDPAP-----PAAAAAALYYSSSTLPAPP 263
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
PP-----LGS-----FADGLDRHYLLRGGGGPAATLSRTYHSS----- 187
LAAPQGGSPKLRQGSAPGATYAAPR-----GSSPKQS-----PSRLAKSYSTSS 315
--GFPGPESRDIPSYGSLSRGLGVPRPTGLLPGPGDGCFTLPCR-REAPMG 242
VSSAGLSPIRVTSPTVOSTISSPIHQLSSTIGTYATLSPTKRLVHASEQYKX 375
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
-----SGPPSGRSL-----PEHFQAEPYG-----LEDDTRSLAADDEGGPD 279
ATATLORPGLSAAAGSRASYSSOHGHLGPELRALQSPHHIDPIYEDRVYQKPPMR 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
STATRRP-----BYGR-GURARAFEDTADGELIER----- 317
QGDPLPFAHTGTTRTSTAPSSPGVDSVPLQRTGSHGQPQNAATAFQASAAAGP 495
-----PP-----FPAATAPLAQ-----PERGSLGLDR----- 340
ADPYRLQVCPVSPSYSGPALPPEGLTARSPSIDSTOKOPREFGWRDPPEV 555
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
-----VVRSPSVDSRKEPR-----WRDPPELPEV 365
HQFPVQSNAAAYLOHLCFGDNKIKABIRROGGIOLLVLLDHRMTVEHRSACGA 615
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
HPVDVKNAAAYLOHLCFENEGIKRRVRQURGLPFLVALLDHPRAEYRRACGA 425
YKANDDNKIALKNCGGIPALVRLRKTTDLRELVGTGLWNLSSCDALXWPII 675
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
YGR-DTDNKAIRDGGVPALVRLRAARDNEVRELVTGTLWNLSSYEPLOXVII 484
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

676 QDALAVLTNAVIIIPHSGWENSPLOQDRKIQLHSSQVLRNATGLRNVSAGEEARR
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:
485 DHGLQTLTHEVIVPHSGWELEPNEDSKPRDAEWTTVFKNTSGCLRNVSDDGAEARR
QY :|||:|||||:|||||:|||||:|||||:|||||:|||||:
736 CDGLTALLVVIQSALGSSEIDSKTVENCVCILRNLSYRLAAETSCQOHMGTD--
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:
545 CEGLVALLHALQSAVGRKDTDNKSVENCVCIMRNLSYHVHKEVPGADRYQEAEEG
QY :|||:|||||:|||||:|||||:|||||:|||||:|||||:
794 LCGEANGKDAESSGCGWKKKKKSDQMD--GUGPLPDCAEPPKGIOMLWHPST
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:
605 TTSQRRKD--DASCFGGKAKGKKDAEMDRNFDTLDPKRTAAKGFELLYQPEV
QY :|||:|||||:|||||:|||||:|||||:|||||:|||||:
851 LTLSECSNPDTLEGAGALQNLAAQSKWVYIRAAVRKKGKGLPILVELLRIDND
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:
663 LSLTSESRNPTLEAAAAGALQNLASGNTWATYIRATVRKERGLPVLVELLQSETD
QY :|||:|||||:|||||:|||||:|||||:|||||:|||||:
911 AVATALRNMALDVNRKELIGKYAMRDIVHRLPGNNSNNTA--SKAMSDDTVTAVC
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:
723 AVAIALRNJSLDQNRKDLIGSYAMTBLVRNV---RNAQAPAHPSAHLEEDTVVAVL
QY :|||:|||||:|||||:|||||:|||||:|||||:|||||:
969 EVITKMNENAKALRDAGGIEKLVGISKSGDKHSPKVVKAASOVLSNMWQYRDLRS
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:
780 EIVSDSLDNARSLQARGVPALVALVAS---SQSVREAKAASHVLQTVWSYKELRG
QY :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1029 DGMSQHFVASSSTIERDRQRPYSSRTSPISPV--RVSPNNRSASAPASPRE---
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:
837 DGWTKGRF-QSASTAKPGKTPSSGGFDSTLPLVDKSLDGEKSNTRDVPMDTLG
QY :|||:|||||:|||||:|||||:|||||:|||||:|||||:
1080 -MISLKERKTDYECTGSNATYHGAKGHESTRKDM 1113
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:
896 ATVDRRERT---LGSDDST---GD-TSEKELL 920

```

RESULT 9

```

Q8C3U7 PRELIMINARY; PRT; 956 AA.
AC Q8C3U7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Armadillo repeat gene deleted in velo-cardio-facial syndrome.
GN ARVCF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Lung;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotati-
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR ENBL; AK084886; BAC39302.1; -.
DR MGD; MGI:109620; Arvcf.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 4.
DR PROSITE; PS0176; ARM_REPEAT; 3.
SQ SEQUENCE 956 AA; 104256 MW; D20921E1D5CB437E CRC64;

```

```

Query Match 22.1%; Score 1411; DB 11; Length 956;
Best Local Similarity 34.7%; Pred.No.7.6e-82;
Matches 386; Conservative 146; Mismatches 335; Indels 246; G;

```

```

QY 47 TTSAILASVKEQLQERLTRELEAREQIVASQERCKL-GSETGSMSSMSAEQI
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

[illegible]

RESULT 10	Q924L5	PRELIMINARY;	PRT;	962 AA.
ID	Q924L5			
AC	Q924L5;			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	ARVCF isoform A1.			
GN	ARVCF.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mu			
OX	NCBI_TaxID=10090;			
RN	[1]			
RS	SEQUENCE FROM N.A.			
RP	STRAIN=C57BL/6;			
RA	Saint-Jore B., Puech A., Merscher S., Xu H., Kuchterlapati R.,			
RA	Skoultschi A.;			
RT	"Developmental expression analysis of Arvcf, a candidate gene fo			
RT	velo-cardio-facial syndrome."			
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF286212; AAK4214.1; -.			
DR	MGD; MGI:109620; Arvcf.			
DR	GO; GO:0005737; C:cytosolasm; IDA.			
DR	GO; GO:0005634; C:nucleus; IDA.			
DR	GO; GO:0005886; C:plasma membrane; IDA.			
DR	GO; GO:0005515; F:protein binding; IPI.			
DR	InterPro; IPR008938; ARM.			
DR	InterPro; IPR000225; Armadillo.			
DR	Pfam; PF00514; Armadillo_seg; 4.			
DR	SMART; SM00185; ARM; 5.			
DR	PROSITE; PS0176; ARM_REPEAT; 3.			
SQ	SEQUENCE 962 AA; 105022 MW; 17B618AB987E1C55 CRC64;			
Query Match	22.1%;	Score 1411;	DB 11;	Length 962;
Best Local Similarity	34.5%;	Pred. No. 7.7e-82;		
Matches	388;	Conservative 148;	Mismatches 325;	Indels 264;
Qy	47	TTSAIIASVKEQLQFELTTRLEAERQIVASOLERCKL-GSETGSMSSMSAAEE		
Db	8	SAASIIASVKEQEARFELTALQERRHVALQLERAQGMSSGGVVG-SGQPL		
Qy	106	SDQGGKDELTGGLVDSICRSLQESGILDPDYSTGERPSLLSQ-----SAL		
Db	67	Q-----LVLQEQSPGQASLATMPEAPEVL		
Qy	162	PGSFQYPASVHSNLTALGETTSPQLPARQTQARATGQSFSGTTSRAGHL-AG		
Db	97	VREDPGTPTSHVSIIVTSEDGTT-----RRTEKVT-KIVKIVTTRTVQVDPGLG		
Qy	221	-----PPPPPPRPPFAPSLGSAFHLDPAP-----PAAAAAALYSYST		
Db	149	LLDGGPP-----LGS---FADGFLDRHYLLRGCGGPAATLSRTYHSS--		
Qy	264	RGSPFLAAPQGGSPTKLORGSAPEGATYAAPR-----GSSPKOS-----PSLAKS		
Db	188	-GG---GFPDGPESRDIPSYGLSRG-GVRRPPTGLLGPFGDGCFTLPGR-REA		
Qy	316	PINIVSSAGLSPIRVTPPTVQSTTSSPIHQLSSSTIGTYATLSPTKRLVHASE		
Db	243	E-----SGPPSGRL-----PEHQAEPYG---LEDDTRSLAADD		
Qy	376	SGELYATATLQPGSLAAGSRASYSSOHGLGPELRAIQSEPHHIDPIVEDRVYQ		
Db	280	LEPDYSTATRRRP-----EYGR-GLRARAFEDTADAGELIEER-----		
Qy	436	SLSQSGDELPFAHTGTVTETSTAPSSFGVDSVPLQRTGSGHQGNAAATFORAS		
Db	318	-----PP-----PFAATAPLAQ-----PERGSLSLDR--		
Qy	496	ASNYADPRLQYCPSPVSPSKSGPALPPEGTILARSFSDTSQKQPRFGFMWRD		


```

RELIMINARY; PRT; 295 AA.
TremBLrel. 19, Created)
TremBLrel. 19, Last sequence update)
TremBLrel. 25, Last annotation update)
(Novel gene similar to human, rat and mouse CTNND2
a 2, cadherin-associated protein delta 2, neural
elated arm-repeat protein) (Fragment).
erio (Zebrafish) (Danio rerio).
tazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
i; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
anio.
55;
N.A.
T-2001) to the EMBL/GenBank/DBJ databases.
7; CAC94831.1; -.
008938; ARM.
00225; Armadillo.
; Armadillo_seg; 3.
5; ARM; 3.
176; ARM_REPEAT; 2.
1
1
95
5 AA; 32352 MW; 995237FAA98E6A09 CRC64;
21.2%; Score 1353.5; DB 13; Length 295;
arity 86.7%; Pred. No. 6.3e-79;
onservative 16; Mismatches 19; Indels 5; Gaps 1;
GVDSPVLPQRTGSHQGPQNAATAAATQFASAAAGPASNYADPYQIQYCPSPVESPYSK 518
GVDSPVLPQRTASQN-----ATGTPPGYASGQSANTYDIYRTLPYCSPVESPYSK 55
LPPEGTLARSPSIDSIQKDPREFQWRDPELPEVITOMLQHPFSPVQSNAAAYLQHL 578
LPPEGTLARSPSIDSIQKDPREFQWRDPELPEVITOMLQHPFSPVQSNAAAYLQHL 115
KIKAEIRROGGIQLLVLLDHRMTEVHRSAGALRNLYGKANDNDKIALKNCGGI 638
KIKAEIRROGGIQLLVLLDHRMTEVHRSAGALRNLYGKANDNDKIALKNCGGI 175
RLRKTTDLREIRLVTVGLNLSLSSCDALKMPEIIQDALAVLTNAVIIIPHSGWENSPL 698
RLRKTTSDVEVRELVTGVLNLSLSSCDALKMPEIIQDALAVLTNAVIIIPHSTWDVSPH 235
KIQHSSQVLRNATGCLRNVSAGAEARRMRRECGLTDALLYVIQSLGSSSEIDS 758
KIQHSSQVLRNATGCLRNVSAGAEARRMRRECGLTDALLYVIQSLGSSSEIDS 295
RELIMINARY; PRT; 892 AA.
TremBLrel. 23, Created)
TremBLrel. 23, Last sequence update)
TremBLrel. 25, Last annotation update)
eat gene deleted in velo-cardio-facial syndrome.
(Mouse).
tazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
heria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
090;
N.A.
6J;
683; PubMed=12466851;
nsortium,
ome Exploration Research Group Phase I & II Team;
the mouse transcriptome based on functional annotation of

```

```

60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK044982; BAC32169.1; -.
DR MGI; 109620; Arvcf.
DR GO; 0005737; C:cytoplasm; IDA.
DR GO; 0005634; C:nucleus; IDA.
DR GO; 0005886; C:plasma membrane; IDA.
DR GO; 0005515; F:protein binding; IPI.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR00225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 4.
DR PROSITE; PSS0176; ARM_REPEAT; 3.
SQ SEQUENCE 892 AA; 97164 MW; 937AA47D0604ED90 CRC64;

Query Match 20.9%; Score 1334; DB 11; Length 892;
Best Local Similarity 37.2%; Pred. No. 6.1e-77;
Matches 352; Conservative 147; Mismatches 300; Indels 148; C

QY 223 PPPPEFPAP-SLGSAPHLDPAPPAALYSSSTLPAPPGGSLAAPQGGSI
DB 2 PAELRQEQSPGSOASLATMPEAP-----EVLEETVVEEDP-----GTF
QY 282 RGSAPGATYAAPRGSPKQSPRLAKSYSTSPINIVSSAGLSPIRVTSPTT
DB 45 SIVTSEDGTT-----RRTETK--VTKTVTTRTVRVQVPLGPDGL-PLDGGPEL-
QY 342 SSSPI--HQLSSITGVTATLSPTKRLVHASEQYKSHSOELYATATLOR-----
DB 96 ADGLDRLHYLLRGGGPAATLSRTHSSGGGPPGPPESRDIPSYGSLSRGLGVRI
QY 388 ----PGS---LAAGRSY--SSQHGHLGPELRALQSPHHIDPIYEDRVYQKPP-
DB 156 LGPGGPGDCTLPGRRAFPMGSESG--PPSGRSL--PEH-----FOAEPY
QY 434 -MRSI--SQSQDQP-LPAAHTGTYTSTAPSPGVDSVPLQRTGSHQGPQNAATAI
DB 202 DTRSLAADDEGGDLEP-----DYSTAT-----RRPEYG-RGLARAI
QY 491 YAAGPASNYADPYQIQYCPSPVESPYSKSGPALPPEGTIA-----RSPSIDSI
DB 244 DDAGELLEERPPF-----PAATAPL-----AQPERGSLGSLDRVVRSPSVDSTF
QY 544 EFGWRDPELPEVITOMLQHPFSPVQSNAAAYLQHLQHPGDNKIKAEIRROGGIQLV
DB 293 ---WRDPELPEVITOMLQHPFSPVQSNAAAYLQHLQHPGDNKIKAEIRROGGIQLV
QY 604 RMTVEHRSAGALRNLYGKANDNDKIALKNCGGI PALVRLRLKTTDLREIRLVTV
DB 350 PRAEVRRACALRNLSYGR--DTDNKAAIRDCGGVPALVRLRLRAARDNEVRELVTG
QY 664 LSSCDALKMPEIIQDALAVLTNAVIIIPHSGWENSPLQDDRTKIQHSSQVLRNATGCI
DB 409 LSSYEPLKWIIDHGLQTLTHEVIVPHSGWEREPNEDSKPRDAEWITVFKNTSGCI
QY 724 SAGEARRMRRECGLTDALLYVIQSLGSEIDSKTVENCVCIIRNLNLSYLAAEI
DB 469 SDGAEARRMRRECGLVDALLHALQSAVGRKDTDNKSVENCVCIMRNLNLSYHVHKEV
QY 784 HMGITDE--LDGLLGEANGKDAESSGCGWKKKKKQDQWD---GVCPLPDCAEI
DB 529 RYQAEAPGIPGSTTSQRRRKD--DASCFGGKAKGKKDAEMDRNFDITLDPRTTEP
QY 839 QMLWHPISVPIYLTLLSECSNPDTLEGAAGALQNLAAAGSMKWSYVIRAAVKEKGL
DB 587 ELLYQPEVRLYLSLLTESRNFNTLEAAGALQNLASAGNTWTATYIRATVRKERGL
QY 899 ELLRIDNDRVVCAVATARNMALDVNRKELIGKYMARDLVHRLPGGNNSNTA--S
DB 647 ELLQSETDKVVRVAIAIRNLNLSLQDRNKLTGSYAMTELVRV---RNAQAPAPHS
QY 957 DDTVTAVCCITLHEVITKNMENAKALRDAGGLEKLVGISKSGDKHSPKVKVKAASQV
DB 704 EDTVAVLNTIHEIVSLSLNDARSLLQARGVPALVALVAS----SQSVREAKAASHV

```

